National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

NL-339 (r. 88/04)

Canada

THE UNIVERSITY OF ALBERTA

Nucleotide Sequence of a *Neurospora* Mitochondrial Plasmid

by

Suchira Pande

©

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF GENETICS

EDMONTON, ALBERTA
FALL 1988

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright-owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-45484-9

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Suchira Pande

TITLE OF THESIS: Nucleotide Sequence of a Neurospora
Mitochondrial Plasmid

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 1988

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permssion.

..........S pande..........

C-7/120, S.D. Area
N. Delhi, 110016
India

Date: April 25, 1988

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Nucleotide Sequence of a Neurospora Mitochondrial Plasmid submitted by Suchira Pande in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Frank Nargang

Frank E. Nargang

Supervisor

John B Bell

John B. Bell

Kenneth L. Roy

Kenneth L. Roy

RBHodgetts

Ross B. Hodgetts

A.J.F.

Anthony J. F. Griffiths

External Examiner

Date: April 25, 1982

I dedicate this thesis to my respected mother, Shrimati Mohini Bala Pande and to the memory of my dear late father Shri Bhagirath Pande. Without their vision of future and belief in me this project would not even have started. The spiritual and the moral support given by Swami Maheshanand Giriji and my family provided me the strength to successfully complete this undertaking.

ABSTRACT

The mitochondrial plasmid found in the Neurospora intermedia strain LaBelle-1b consists of 4,070 base pairs and is relatively AT rich (65% A+T). Translation of the DNA sequence using the Neurospora mitochondrial genetic code reveals the presence of an open reading frame that is 3,456 nucleotides long. This long open reading frame could encode a soluble, hydrophilic, relatively basic protein composed of 1,151 amino acids with a molecular weight of 131,727 daltons. The plasmid is transcribed such that the long open reading frame could be expressed. The codon usage of this long open reading frame indicates that the LaBelle plasmid may be related to mitochondrial introns. Based on its similarity to other proteins it is conceivable that this 1,151 amino acid long polypeptide encodes an RNA polymerase.

A region of 1,085 base pairs of the LaBelle plasmid contains homology to a region of 1,310 base pairs of the EcoRI-6 fragment of N. intermedia LaBelle-1b mtDNA. This homologous region of the EcoRI-6 fragment is almost totally conserved between the N. intermedia strain LaBelle-1b and Neurospora crassa 74A mtDNA. Since the LaBelle plasmid sequence contains much greater divergence than the two mitochondrial DNAs, it is possible that the plasmid existed before the separation of the two species and that divergence from the parent sequence has been occurring for some time. Alternatively, it is possible that N. intermedia acquired the plasmid after the species separated. If this were the case, the region of homology on the LaBelle plasmid

might represent an example of convergent evolution. This is an attractive hypothesis in view of the possibility that the region of homology may contain an origin of replication. That is, a foreign genetic element might have increased its fitness within mitochondria by evolving a replication origin similar to that of the host mitochondrial DNAs.

ACKNOWLEDGEMENTS

I am grateful to Dr. Frank Nargang, my Ph.D. supervisor, for all the instruction, advice, support, patience and help that he gave me throughout the course of this study. To Dr. Ken Roy, I express my gratitude for his support and stimulating discussions. The attention that Dr. Ross Hodgetts gave me and the enthusiasm that he generated in me have been of tremendous help during all these years of graduate studies. The interest and enthusiasm of Dr. John Bell to know more about India and her culture made it a pleasure to share my heritage with him. I am grateful to Dr. Wayne Anderson for his critical comments on parts of this thesis.

To all the members of Dr. Nargang's laboratory, thank you for introducing me to the various aspects of Canadian culture and teaching me to appreciate the things in life that I had always taken for granted.

I would like to thank Mariola Drygas for being more like a sister here in Edmonton than a mere friend. Edmond Lemire, you made it easier to go through difficult times of the graduate program by just being yourself. Drell Bottorff, but for your help getting the various sequence printouts for this thesis would have been a nightmare. I am grateful to Dr. Avalokitesvara Sen for the advice and help that he gave me in preparing the graphics for this thesis. The helpful suggestions and the technical assistance provided by Mr. William Clark throughout this project in numerous ways was invaluable. Special thanks are due to Mr. Perry d'Obrenan who was very helpful in densitometric scanning of the negatives. Last, but not the least,

thanks to all the friends that I made here in Edmonton for enlightening me about their countries and cultures.

This study was supported by the Teaching Assistantship provided by the Department of Genetics in 1982-1983 and the Alberta Heritage Foundation for Medical Research (AHFMR) Studentship from 1983-1988.

TABLE OF CONTENTS

CHAPTER PAGE

I INTRODUCTION.................................... 1

1. Establishing the location of *Neurospora* mitochondrial plasmids........................ 4

2. Size estimation of the *Neurospora* mitochondrial plasmids........................ 6

3. *Neurospora* homology group 1 mitochondrial plasmids.................................... 7

4. *Neurospora* homology group 2 mitochondrial plasmids.................................... 8

5. *Neurospora* homology group 3 mitochondrial plasmids.................................... 14

6. *Neurospora* nuclear plasmids.................. 14

7. Vectors for *Neurospora* transformation........ 17

II MATERIALS AND METHODS........................... 20

1. Strains of *Neurospora*........................... 20

2. Strains of *E. coli*................................ 20

3. Recombinant plasmids.............................. 20

4. Maintenance of the strains........................ 21

5. Growth media...................................... 22
   - i) Vogel's medium............................... 22
   - ii) L-broth...................................... 22
   - iii) Minimal medium.............................. 22
   - iv) X-Gal plates................................. 22

6. Bacterial transformation.......................... 22

7. Large scale plasmid DNA preparation............... 23

8. Miniprep procedure for plasmid isolation.......... 25

9. Transfection...................................... 26

10. Isolation of single stranded DNA.................. 27

TABLE OF CONTENTS

CHAPTER PAGE

11. Restriction enzyme analysis........................ 27

12. Agarose gel electrophoresis of DNA................ 27

13. DNA recovery from agarose gel........................ 28
    i) Recovery of DNA from LMP gels.............. 28
    ii) Recovery of DNA using DE81 paper........... 29
    iii) Recovery of DNA using NA-45 membrane...... 29

14. Ligations.. 30

15. Radiolabeling of DNA................................ 30

16. Isolation of mitochondria........................... 31

17. Isolation of mitochondrial plasmid DNA............ 32

18. Cytochrome spectra................................... 33

19. DNA sequencing... 33

20. DNA sequencing gels.................................. 36

21. DNA sequence analysis................................ 36

22. Isolation of RNA from mitochondria................ 37

23. Preparation of labelled riboprobes................. 38

24. RNA electrophoresis.................................. 39

25. Northern blotting..................................... 40

26. RNA-RNA hybridization................................ 40

27. Autoradiography....................................... 41

III RESULTS AND DISCUSSION................................ 42

I Analysis of the LaBelle plasmid.................... 42

1. DNA sequence of the LaBelle mitochondrial plasmid.. 42

TABLE OF CONTENTS

CHAPTER PAGE

2. G+C rich regions in the LaBelle plasmid...... 42
3. The LaBelle plasmid contains a long open reading frame........................... 43
4. Codon usage in the long open reading frame.. 46
5. Absence of conserved sequence elements of mitochondrial introns...................... 48
6. Palindromic sequences........................ 49
7. Direct repeats............................... 51
8. Comparison of the LaBelle plasmid DNA sequence to the DNA sequences in the Gene Bank database................................ 53
9. Comparison of the deduced amino acid sequences coded by the reading frames found in the LaBelle plasmid with protein sequences in the Gene Bank database........................... 58
10. Repeated motif in the 1,151 amino acid long polypeptide................................ 66
11. Transcription of the LaBelle plasmid......... 66
12. Sizes of the LaBelle transcripts............. 71

II DNA sequence analysis of homologous regions in Neurospora mtDNA and LaBelle mitochondrial plasmid DNA..................................... 75
1. PstI palindromes............................. 78

III Heterokaryon analysis............................ 80

IV CONCLUSIONS.. 82
1. The LaBelle plasmid.......................... 82
2. Region of homology to mtDNA.................. 87

TABLE OF CONTENTS

CHAPTER **PAGE**

TABLES.. 92

FIGURES... 113

PHOTOGRAPHIC PLATES... 176

BIBLIOGRAPHY.. 185

LIST OF TABLES

TABLE PAGE

1. Plasmids found in Neurospora species.................... 93

2. Base composition of Neurospora genomic DNA, mtDNA and mitochondrial plasmids.................................. 97

3. GC rich regions in the LaBelle mitochondrial plasmid...... 98

4. Positions, lengths, polarity indices, molecular weights, G+C and A+T precentages, and percentage of codons ending in G/C and A/T of ORFs ending at stop codon 3512-3514 and starting at all possible preceeding AUG codons as initiating codons.. 99

5. Positions and lengths of open reading frames (ORFs) in the LaBelle mitochondrial plasmid........................ 101

6. Codon usage in the long (1,151 amino acid) open reading frame of the LaBelle mitochondrial plasmid.............. 102

7. Percentage of codons ending in G or C and A or U......... 106

8. Usage of distinctive codons.............................. 107

9. List of palindromic sequences in the LaBelle mitochondrial plasmid..................................... 108

10. List of direct repeats in the LaBelle mitochondrial plasmid.. 109

11. List of regions in the 1,151 amino acid polypeptide of the LaBelle mitochondrial plasmid that are homologous to regions of various RNA polymerases. The percentage identity and number of correct matches in the regions under consideration are shown........................... 110

12. Insertions present in the N. intermedia EcoRI-6 mtDNA fragment as compared to the homologous region of the LaBelle mitochondrial plasmid (or vice versa)............ 111

13. List of palindromic sequences found in the region of homology between the N. intermedia EcoRI-6 mtDNA fragment and the LaBelle mitochondrial plasmid.................... 112

LIST OF FIGURES

FIGURE PAGE

1. Sequencing strategy used for the LaBelle mitochondrial plasmid.......... 114

2. DNA sequence of the LaBelle mitochondrial plasmid.......... 116

3. Circular restriction map of the LaBelle mitochondrial plasmid 122

4. Positions and lengths of open reading frames (ORFs) in the LaBelle mitochondrial plasmid.......... 124

5. Amino acid sequence of the polypeptide potentially encoded by the longest open reading frame using the *Neurospora* mitochondrial genetic code.......... 126

6. Hydropathy plot of the putative 1,151 amino acid polypeptide. 131

7. Palindromic sequences in the LaBelle mitochondrial plasmid with free energies of formation (ΔG) equal to or less than -10 Kcal/mol.......... 133

8. Repeated regions and consensus sequences; their similarities to the LaBelle mitochondrial plasmid and the *N. intermedia* EcoRI-6 mtDNA sequences.......... 135

9. Regions of identity between the putative 1,151 amino acid polypeptide of the LaBelle mitochondrial plasmid and various RNA polymerases.......... 137

10. Sequence of the DNA binding domain of λ repressor, λ Cro and the putative protein encoded by the LaBelle long ORF.......... 140

11. Recombinant plasmids pBSA1 and pBSA15 constructed to prepare single stranded riboprobes and cRNA.......... 142

12, Map of N. cra*ssa* mtDNA.......... 144

13. Linear map of the EcoRI-6 fragment of Neurospora mtDNA showing the region homologous to the LaBelle mitochondrial plasmid DNA.......... 146

14. Sequencing strategy used for the EcoRI-6 fragment of *N. intermedia* strain LaBelle-1b mtDNA.......... 148

15. DNA sequence of a portion of the EcoRI-6 fragment of *N. intermedia* strain LaBelle-1b mtDNA.......... 150

LIST OF FIGURES

FIGURE **PAGE**

16. Nucleotide sequence of homologous regions between LaBelle mitochondrial plasmid and *N. intermedia* EcoRI-6 mtDNA fragment.. 153

17. Analysis of the region of homology between the LaBelle mitochondrial plasmid and the *N. intermedia* EcoRI-6 fragment. 165

18. Palindromic sequences capable of forming hairpin loop structures with free energies of formation (ΔG) equal to or less than -10 Kcal/mol.in the EcoRI-6 fragment of *N. intermedia* mtDNA.. 167

19. Sequencing strategy used for the EcoRI-6 fragment of *N. crassa* mtDNA.. 169

20. DNA sequence of a portion of the *N. crassa* EcoRI-6 mtDNA fragment.. 171

21. Potential hairpin loop structures in the region of homology between the LaBelle mitochondrial plasmid and the *N. intermedia* EcoRI-6 mtDNA fragment............................. 174

LIST OF PHOTOGRAPHIC PLATES

PLATES	DESCRIPTION	PAGE
I	Northern blot showing hybridization of various probes to mtRNA isolated from the LaBelle plasmid-containing strain	177
II	Ethidium bromide stained 1.5% agarose formaldehyde gel showing RNase A digestion is effective....................	179
III	Northern blot performed to assess the quality of mtRNA preparation..	181
IV	Northern blot showing LaBelle mitochondrial plasmid specific transcripts.....................................	183

LIST OF ABBREVIATIONS

bp	Base pairs
bromophenol blue	3', 3", 5', 5"-tetrabromophenolsulfonphthalein
BSA	Bovine serum albumin
cpm	Counts per minute
DEPC	Diethylpyrocarbonate
dH_2O	Distilled water
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid, disodium salt
EGTA	Ethyleneglycol bis (β-Amino-Ethyl-Ether) N, N, N, N- tetraacetic acid
EPPS	N-[2-Hydroxyethyl] piperazine-N' -3-propanesulfonic Acid
FGSC	Fungal Genetics Stock Center
HEPES	N-[2-Hydroxyethyl] piperazine-N' -2-ethanesulfonic Acid
IPTG	Isopropylthio-β-D-galactoside
Kb	Kilobase
ORF	Open reading frame
PEG	Polyethylene glycol
RNA	Ribonucleic Acid
RNase	Ribonuclease
RNasin	Ribonuclease inhibitor
rRNA	Ribosomal Ribonucleic Acid
rpm	Revolutions per minute

LIST OF ABBREVIATIONS

SDS	Sodium dodecylsulfate
TEMED	N, N, N', N' - Tetramethylethylenediamine
Tris	Tris (hydroxmethyl) aminomethane
Triton X-100	Octyl phenoxy polyethoxyethanol
tRNA	Transfer Ribonucleic Acid
URF	Unidentified reading frame
UV	Ultraviolet
X-Gal	5-Bromo-4-chloro-3-indolyl-β D-galactoside

CHAPTER I

INTRODUCTION

Bacterial plasmids were discovered during early studies of bacterial recombination (Campbell, 1969). During the 1960's they attracted attention when plasmids were shown to confer antibiotic resistance to their host (for review see Foster, 1983). The first plasmid found in eukaryotes was the 2-μm DNA circle of *Saccharomyces cerevisiae* (Sinclair *et al*., 1967). For almost a decade the existence of plasmids in eukaryotes was considered unique to yeast. The discovery of plasmids in the mitochondria of *Zea mays* (Pring *et al*., 1977) extended this domain to higher plants. A short time later, Stahl *et al*., (1978) and Cummings *et al*., (1978) independently demonstrated the existence of plasmids in the mitochondria of the filamentous fungus *Podospora anserina*.

There is no generally agreed to definition of mitochondrial (mt) plasmids, and the different definitions used by different investigators have created some confusion in the literature. For example, the term mitochondrial plasmid has been used in the literature to include molecules which are in fact defective versions of the normal mitochondrial genome of the host organism (Esser *et al*., 1983). These are derived wholly from some portion of the parent mtDNA, have the capacity to replicate independently, and are generally suppressive toward the parental mtDNA. Ragged mutants of *Aspergillus* (Lazarus *et al*., 1980), stopper mutants of *Neurospora* (Bertrand *et al*., 1980), petite mutants of yeast (Locker *et al*., 1979) and senescing strains of *Podospora* (Jamet-Vierny *et al*., 1980; Wright *et al*., 1982) are examples where the term mitochondrial plasmid was used in this sense. Defective mtDNAs are not always suppressive relative to wild-type mtDNA, as in the case of *Cochliobolus heterostrophus* (Garber *et al*., 1984). The latter small circular DNAs are not suppressive, but have been found to consist of mtDNA segments that contain the gene for ATPase subunit VI (Garber *et al*., 1987).

Tudzynski, (1982) defined "true" mitochondrial plasmids as small circular or linear DNA species that are present in mitochondria, but which are not derived from the standard mtDNA. This definition would exclude the defective mtDNAs listed above. Such plasmids have been found in the mitochondria of many filamentous fungi such as *Claviceps purpurea* (Tudzynski *et al*., 1983; Tudzynski and Esser, 1986), *Ascobolus immersus* (Francou, 1981), *Neurospora* (Nargang, 1985), *Agaricus bitorquis* (Mohan *et al*., 1984), and *Fusarium oxysporum* f. sp. *conglutinans* (Kistler and Leong, 1986).

Lambowitz *et al*., (1987) define mitochondrial plasmids as DNA or RNA species that are present in mitochondria, but are not derived from the standard mtDNA (though they may be homologous to the standard mtDNA for the reasons given below). They include all the nucleic acid molecules that originated independently of mtDNA and possibly outside of mitochondria, in the nucleus or cytosol, or as infectious elements from other organisms as mitochondrial plasmids. By emphasizing the evolutionary origin of the plasmids it is possible to consider excised forms of mtDNA introns and some other sequences as mitochondrial plasmids since these may themselves have originated by insertion of foreign elements into mtDNA, as mitochondrial plasmids. Thus α sen DNA of *Podospora anserina* (Osiewacz and Esser, 1984) and possibly the *Kalilo* element of *N. intermedia* (Bertrand *et al*., 1985) could be considered as mitochondrial plasmids.

In the present work the term mitochondrial plasmid will be used only for the so-called "true" mitochondrial plasmids. The α sen DNA and other sen DNAs of *Podospora* will be referred to as excision-amplification plasmids (Cummings *et al.*, 1987) and the *Kalilo* element will be referred to as a nuclear plasmid (Bertrand, 1987).

Standard laboratory strains of *Neurospora* do not contain mitochondrial plasmids. They were discovered accidentally during searches for mtDNA polymorphisms in strains recently isolated from nature (Collins *et al.*, 1981). Since then, several mitochondrial plasmids have been found in different species of *Neurospora* (Stohl *et al.*, 1982; Natvig *et al.*, 1984; Taylor *et al.*, 1985). The presence of mitochondrial plasmids in *Neurospora* strains does not impart a readily detectable phenotypic characteristic which could be attributed to the presence of the plasmids. The cytochrome spectra and mtrRNA profiles of the plasmid-containing strains are also indistinguishable from those of the wild-type strains (Collins *et al.*, 1981; Stohl *et al.*, 1982). So far, it is not known whether these mitochondrial plasmids of *Neurospora* have some biological function or whether they represent "selfish DNA".

Initially three different mitochondrial plasmids were identified in three different wild-type *Neurospora* strains: *N. crassa* strain Mauriceville-1c (Collins *et al.*, 1981) and *N. intermedia* strains LaBelle-1b and Fiji (Stohl *et al.*, 1982). These three mitochondrial plasmids are different from each other based on DNA-DNA hybridization studies. Subsequently, an additional member of the Mauriceville homology group was identifed in an *N. intermedia* strain from Varkud, India (Akins _et_ _al_., Submitted). Natvig _et_ _al_., (1984) identified four additional mitochondrial plasmids, three from Hawaii and one from Surinam, all in _N_. _tetrasperma_. All four of these were shown to be homologous to the Fiji mitochondrial plasmid from _N_. _intermedia_ (Natvig _et_ _al_., 1984). Based on DNA-DNA hybridization data Natvig _et_ _al_., (1984) grouped the mitochondrial plasmids into three different homology groups. With the identification of the plasmid in the _N_. _crassa_ Roanoke strain and the demonstration that the Roanoke mitochondrial plasmid is homologous to the Fiji (from _N_. _intermedia_) and Hanalei (from _N_. _tetrasperma_) mitochondrial plasmids of homology group 1, it became clear that similar plasmids are found in different species (Taylor _et_ _al_., 1985). Furthermore, it is now known that strains containing more than one mitochondrial plasmid also exist (S. Vollmer pers. comm. in Taylor _et_ _al_., 1985). The known _Neurospora_ plasmids are listed in Table 1.

None of these _Neurospora_ mitochondrial plasmids have been found integrated in mtDNA in wild-type strains, nor have there been reports that plasmid sequences are integrated in nuclear DNA. However Akins _et_ _al_., (1986) and Lambowitz _et_ _al_., (1987) have isolated _Neurospora_ mitochondrial mutants that contain the Mauriceville and the Varkud plasmid sequences integrated into the mtDNA. These were isolated as growth impaired mutants of the plasmid-containing strains under conditions favouring the accumulation of suppressive DNAs in the mitochondria (discussed more fully in section 4 below).

(1) Establishing the location of _Neurospora_ mitochondrial plasmids

Different approaches were taken to establish that various

Neurospora plasmids were located in the mitochondria. DNA was isolated from mitochondrial preparations that were highly purified by flotation gradient centrifugation. Upon restriction enzyme digestion of whole mtDNA so prepared, plasmid DNA could be identified as intense bands on ethidium bromide-stained agarose gels. Micrococcal nuclease treated mitochondria of the Mauriceville, Fiji, and LaBelle strains had both mtDNA and plasmid DNA protected from nuclease digestion whereas closed circular pBR322 DNA, added to the mitochondrial preparation, was completely digested (Collins *et al*., 1981; Stohl *et al*., 1982). Treatment of Mauriceville mitochondrial preparations with increasing concentration of Triton X-100 caused both mtDNA and plasmid DNA to have a parallel increase in sensitivity to micrococcal nuclease. Since DNA in intact λ phage particles was not digested by micrococcal nuclease under similar conditions of Triton X-100 treatment, these results indicated that the Mauriceville mitochondrial plasmid DNA was protected from the micrococcal nuclease digestion by mitochondrial membranes rather than by bound proteins (Collins *et al*., 1981). Cell fractionation experiments clearly showed that the Fiji plasmid is intramitochondrial in location (Stohl *et al*., 1982). However, cell fractionation experiments in which plasmid and mtDNA sequences were assayed in different cell fractions of the LaBelle strain showed that the post-mitochondrial supernatant contained a small amount of the EcoRI fragment corresponding to plasmid DNA, but no mtDNA EcoRI fragments, even after extended exposure of the autoradiograms. The presence of LaBelle plasmid in the post-mitochondrial supernatant was postulated to be due to preferential leakage of plasmid DNA from the mitochondria. However, it was also suggested that some proportion of the LaBelle plasmid may be located in the cytosol (Stohl *et al*., 1982). Reciprocal crosses involving the Mauriceville and the LaBelle strains show strict maternal inheritance of the Mauriceville and the LaBelle mitochondrial plasmids. In sexual crosses the Fiji strain was inefficient in the formation of protoperithecia and, hence, was used only as a conidial (male) parent. In such crosses, none of the progeny contained the Fiji plasmid. The mtDNA also showed strict maternal inheritance in these crosses. Thus, the inheritance patterns of the Mauriceville, Fiji and the LaBelle plasmids are consistent with the intramitochondrial location established by other experiments (Collins *et al*., 1981; Stohl *et al*., 1982).

(2) Size estimation of the *Neurospora* mitochondrial plasmids

The sizes of all the mitochondrial plasmids were estimated by restriction digest analysis (Table 1). These estimates were further supported in the case of the Mauriceville, LaBelle, and Fiji plasmids by electron microscopic (EM) observations of CsCl-ethidium bromide gradient purified plasmid DNA. The monomer contour length of the Mauriceville plasmid was estimated to be 3.6 ± 0.22 Kb, that of the LaBelle plasmid was estimated to be 4.3 Kb and that of the Fiji plasmid was estimated to be 5.3 Kb by EM observations compared to the 3.6 Kb, 4.1 Kb and 5.2 Kb lengths respectively, as estimated by restriction analysis. These monomers also exist in oligomers joined in a head to tail orientation. Oligomers of up to 6 tandemly joined monomers have been measured for all the above plasmids in EM studies (Collins *et al*., 1981; Stohl *et al*., 1982).

(3) ***Neurospora*** **homology group 1 mitochondrial plasmids**

The first member of homology group 1 mitochondrial plasmids was discovered in *N. intermedia* Fiji N6-6 (FGSC #435) by Stohl *et al*., (1982). No prominent detectable transcript was observed for the Fiji mitochondrial plasmid (Stohl *et al*., 1982). Studies by Natvig *et al*., (1984) revealed the presence of mitochondrial plasmids in Hawaiian isolates of *N. tetrasperma* (Table 1). By restriction analysis the plasmids of the Hawaiian isolates (Lihue, Hanalei and Waimea Falls) were indistinguishable from each other and they differed only slightly (by one restriction site) from the *N. tetrasperma* Surinam plasmid. On the other hand the *N. intermedia* Fiji plasmid, though clearly related to these *N. tetrasperma* plasmids as judged by restriction maps, differed substantially (by 4 restriction sites) from the Hawaiian isolates and by 3 such sites from the Surinam isolate (Taylor *et al*., 1985). Natvig *et al*., (1984) put forward two hypotheses concerning the distribution and origin of these plasmids: (1) The group 1 plasmid might have been present before the divergence of *N. intermedia* and *N. tetrasperma* and has simply been retained in both species over time while undergoing evolutionary divergence; and (2) The plasmid might have been transferred between the species since their divergence by interspecific crossing, hyphal fusions, or some form of infectious transmittance.

A circular mitochondrial plasmid present in *N. crassa* 516 (Roanoke, LA) was shown to be identical to the Surinam plasmid from *N. tetrasperma*. It differed by one restriction site from the Hanalei plasmid and 3 restriction sites from the Fiji plasmid of homology group 1 (Table 1)(Taylor *et al*., 1985). Out of 14 Louisiana strains examined; half, including 5 Roanoke isolates and one each from Houma and Welsh contained plasmids homologous to the Roanoke 516 plasmid. By restriction endonuclease analysis, it was determined that these plasmid-bearing Roanoke isolates exhibit 3 different mtDNA types, which were duplicated in plasmid-free isolates from Houma or Welsh (Taylor *et al.*, 1985). This kind of plasmid distribution in isolates with identical mitochondrial genomes was thought to be 1) due to the loss of plasmids from plasmid-bearing individuals or 2) due to the spread of plasmids from plasmid-bearing isolates to plasmid-free isolates independently of mitochondrial inheritance (Taylor *et al.*, 1985).

This latter possibility was experimentally tested and supported by analyzing the progeny of matings between plasmid-bearing conidiating parents and plasmid-free protoperitheciating parents. Six percent of the progeny of such matings were shown to inherit plasmid DNA. These data are in contrast to the strict maternal inheritance of the Mauriceville and LaBelle mitochondrial plasmids in reciprocal crosses (Collins *et al.*, 1981; Stohl *et al.*, 1982). However, it should be noted that this paternally inherited plasmid DNA was not found in progeny mitochondria, and, because it was unstably inherited during vegetative growth, it was suggested to be freely located in the cytoplasm or in the nucleus (Taylor and May, 1987).

(4) *Neurospora* homology group 2 mitochondrial plasmids

The entire DNA sequence of a group 2 mitochondrial plasmid from the Mauriceville-1c strain of *N. crassa* has been determined (Nargang *et al.*, 1984). Analysis of the sequence revealed the presence of a long open reading frame (ORF). This long ORF could encode a polypeptide as large as 710 amino acids (82,000 daltons). The predicted polypeptide was relatively hydrophilic and basic in character (Nargang *et al*., 1984). Determination of codon usage in this long ORF showed: 1) the Mauriceville plasmid was less biased against codons ending in C or G than other fungal mtDNA genes, 2) the codon usage in this plasmid was similar to that in intronic ORFs in fungal mtDNA, and 3) the codon usage in this plasmid differed from that in *Neurospora* nuclear genes (Nargang *et al*., 1984). The Mauriceville plasmid was found to be transcribed *in vivo* to give abundant approximately full length transcripts (Collins *et al*., 1981). The major 5' and 3' ends of this transcript were found to be adjacent and slightly overlapping (Nargang *et al*., 1984). The Mauriceville plasmid DNA sequence also contains a set of sequence elements that correspond to the consensus sequence elements characteristic of group I mtDNA introns. These elements were expressed and also correctly oriented with respect to the 5' and 3' RNA termini in the major transcript (Nargang *et al*., 1984). The codon usage in the ORF and the presence of conserved sequence elements in the plasmid sequence indicated that the Mauriceville plasmid was related to group I mtDNA introns. It was suggested that the Mauriceville plasmid may belong to a class of genetic elements that were or are the progenitors of mitochondrial introns, or it may be an excised intron that has acquired the capacity for independent replication and expression in mitochondria (Nargang *et al*., 1984).

No close homologs of conserved sequences characteristic of the group II mitochondrial introns were detected in the Mauriceville plasmid (Nargang *et al*., 1984). However, the long ORF of the plasmid was initially noted to have a short region of homology to a family of reverse transcriptases (Nargang _et_ _al_., 1984). Lang _et_ _al_., (1985) subsequently showed that the protein potentially encoded by the long ORF in the Mauriceville plasmid could be grouped with the putative proteins specified by the long ORFs encoded by four group II mitochondrial introns. On the basis of seven short blocks of amino-acid stretches found to be conserved in the plasmid's long ORF, the four group II intron ORFs and a family of RNA-dependent DNA polymerases, Michel and Lang, (1985) concluded that the putative 710 amino acid long polypeptide encoded by the Mauriceville plasmid was structurally related to the reverse transcriptases of retroviruses. This long ORF sequence of Mauriceville is almost completely conserved in another mitochondrial plasmid found in mitochondria of _N_. _intermedia_ Varkud-1c from India (Nargang, 1986). Though no function has been assigned to the putative products of these ORFs, the conservation of the ORF sequence in two plasmids found in two different species of _Neurospora_ isolated from geographically remote locations (Mauriceville, Texas and Varkud, India) by inference would be a strong argument for conservation of function (Nargang, 1986). It was speculated that the long ORF of Mauriceville might encode a protein required for propagation or expression of the plasmid (Nargang _et_ _al_., 1984). Subsequently, mitochondrial mutants were isolated from both the Mauriceville and the Varkud plasmid-containing strains, all of which contained aberrant plasmids with small insertions and some of which contained insertions of the plasmid sequences into the mtDNA. The analysis of these integration junction points led to the suggestion that these integration events were promoted by site specific DNA endonucleases which were possibly the proteins encoded by the

Mauriceville and the Varkud plasmids (Akins _et al_., 1986; Lambowitz _et al_., 1987). Thus, these mutants seem to support the above speculation that Mauriceville encodes a protein required for propagation and expression of the plasmid.

To investigate whether the Mauriceville and Varkud mitochondrial plasmids could indeed function as mobile elements, as suggested earlier (Nargang _et al_., 1984), the Mauriceville and Varkud strains were grown continuously by vegetative propagation, and growth impaired mitochondrial mutants were isolated (Akins _et al_., 1986). All these mutants contained mitochondrial plasmids that were suppressive relative to mtDNA. These suppressive plasmids were either larger or smaller than wild-type plasmids and most of the variant plasmids contained approximately 75-150 bp insertions of mtDNA. In 8 out of 9 cases, this insertion was found to correspond to or include a mitochondrial tRNA sequence. In all the cases analyzed this inserted mitochondrial tRNA was oriented head-to-tail with respect to the major plasmid transcript and was inserted in the plasmid immediately downstream of the major 3' end of the plasmid transcript (Akins _et al_., 1986). Ten of these mutants contained defective mtDNAs. These defective mtDNAs 1) generally retained all or part of the region that includes the rRNA genes and most mitochondrial tRNA genes (EcoRI-1[1], 4 and 6); 2) are missing or deficient in EcoRI-2 and 3) show novel mtDNA restriction fragments, some of which reflected recombination between the mitochondrial plasmid and mtDNA (Akins et al., 1986). Time course experiments indicated that the primary event was the accumulation of variant plasmid and/or its integration into mtDNA followed by further mtDNA rearrangements. Three of these mutants contained defective mtDNAs caused by integration of plasmids. The integrated plasmids were preceeded by a run of T residues and the plasmid sequence at the plasmid-mtDNA boundary in two of the three cases corresponded precisely to the 5' end of the major transcript of the wild-type plasmids, and, in the remaining case, it corresponded to the end of a small mtDNA insert that was linked to the normal 5' end of the plasmid RNA (Akins et al., 1986). The common feature of these plasmid-mtDNA boundary sequences was that both correspond to major 5' RNA ends. These results could be explained in two ways : 1) the integrated plasmid sequences also passed through an RNA intermediate or 2) the transposition intermediate was a single-stranded DNA primed by RNA synthesized by the plasmid promoters (Akins et al., 1986). The structural similarity noted between reverse transcriptases and the putative 710 amino acid long polypeptide encoded by the Mauriceville plasmid (Nargang et al., 1984; Michel and Lang, 1985), along with the structure of these suppressive mitochondrial plasmids and defective mtDNA molecules, suggests that an RNA intermediate and a reverse transcription step was involved in the generation of these molecules (Akins et al., 1986).

---

[1] The _Neurospora_ mtDNA is digested with EcoRI and resulting fragments are resolved on an agarose gel. The various fragments are numbered from 1 to 11 based on their sizes. The largest fragment is numbered as EcoRI-1 and the smallest fragment is EcoRI-11.

Since all the integration events occurred in the same region of mtDNA which was near the site of deletion in a stopper mutant, E35, of Neurospora (de Vries et al., 1986), the suggestion was made that both intra- and inter-molecular recombination events might be favoured by the structure and/or accessibility of this mtDNA region (Akins *et al.*, 1986). Further, a common core sequence of the 5 nucleotides AGGAG was observed at the integration sites. It was suggested that recombination was promoted by a site-specific double strand endonuclease activity similar to that encoded by the yeast ω^+ intron of the large rRNA gene (Akins *et al.*, 1986). Thus, it is possible that the long ORFs of the Mauriceville and Varkud plasmids may be contributing the reverse transcriptase and DNA endonuclease activities involved in generating these suppressive and integrative plasmids (Akins *et al.*, 1986).

In the case of *P. anserina*, the α-sen DNA, which is involved in the senescence of this fungus, was shown to correspond precisely to a group II intron of the cytochrome *c* oxidase subunit I (COI) gene of this fungus (Osiewacz and Esser, 1984). In this case too, it was speculated that these α-sen DNA molecules were formed as a result of reverse transcription of the circular excised intronic RNA (Osiewacz and Esser, 1984). Michel and Lang, (1985) had observed a reverse transcriptase-like protein encoded by the ORF in α-sen DNA. Though a direct role of this reverse transcriptase activity in formation of α-sen DNA and senescence has not yet been demonstrated, Steinhilber and Cummings (1986) have reported the presence of a RNA-dependent DNA polymerase activity in middle-aged mycelia of *P. anserina* from crude nuclear/mitochondrial extracts. This activity was not present in young non-senescent mycelia.

Another unique feature of the Mauriceville plasmid amongst the known *Neurospora* plasmids is the presence of a PstI cluster in a 400 bp region of the plasmid DNA. DNA sequence analysis by Nargang *et al.*, (1983, 1984) shows the cluster consists of eight PstI sites. These are organized in five palindromic elements resembling the PstI palindromes characteristic of Neurospora mtDNA (Yin et al., 1981; Yin et al., 1982). This cluster of PstI palindromes occurs about 120 bp downstream from the termination codon of the long ORF (Nargang et al., 1983, 1984). The two outer palindromes of the cluster have the canonical sequence for Neurospora mtDNA PstI palindromes (Nargang et al., 1983). Yin et al., (1981, 1982) have shown that PstI palindromes of mtDNA are frequently found at the boundaries of rRNA, tRNA, and mitochondrial protein genes, though their function if any is unknown.

(5) **Neurospora homology group 3 mitochondrial plasmids**

The sole known plasmid in this group is the plasmid from N. intermedia strain LaBelle-1b (P405). Southern hybridization experiments indicated that this plasmid has no sequence homology to any other mitochondrial plasmid of Neurospora. In very long exposures, weak hybridization was detected between the cloned LaBelle plasmid and the mtDNA EcoRI-6 fragment (Stohl et al., 1982). Unlike the Mauriceville plasmid, Northern analysis indicated that the LaBelle plasmid does not make prominent detectable transcripts. Stohl et al., (1982) did observe weak hybridization to heterogenous RNAs having lengths of 2 to 4 kb but the significance of these weak hybridizations was unclear.

(6) **Neurospora nuclear plasmids**

Another, distinct kind of plasmid named "Kalilo" has been found in N. intermedia strains collected from Hawaii (see Table 1 for a complete list of strains containing Kalilo). The Hawaiian word Kalilo means "dying, at death's door, to be hovering between life and death" (Pukui and Elbert, 1957). This name suitably describes the strains containing this plasmid since Kalilo strains undergo a process of clonal deterioration and die after a limited number of transfers, usually less than twenty (Griffiths and Bertrand, 1984). The process leading to vegetative death in Kalilo strains has been defined as "senescence". Characteristic features of senescence are 1) the appearance of an abnormal stop-start growth phenotype in the mycelium (Rieck et al., 1982) and 2) progressive deterioration of mitochondrial functions, including a gradual disappearance of cytochromes $\underline{aa}_3$ and b and an accumulation of rearranged mtDNA (Bertrand et al., 1985). Large insertions, deletions and inversions of mtDNAs are seen in the senescing Kalilo cultures (Bertrand et al., 1985).

In the Kalilo strains of N. intermedia, the initiation of senescence is caused by the insertion of an extramitochondrial plasmid into mtDNA (Bertrand et al., 1986). This extramitochondrial plasmid is a 9.0 Kb linear DNA with 1260 bp inverted terminal repeats at either end. The 5' ends of the molecule are protected by covalently bound protein (Bertrand, 1987; and personal communication). This 9.0 Kb linear plasmid is found closely associated with the nuclei of both the presenescent and senescent Kalilo strains in high copy numbers, but is not present in cells of normal strains which do not undergo senescence. Paradoxically, this nuclear plasmid is maternally inherited (Bertrand et al., 1986). The autonomously replicating, free Kalilo plasmid has not been detected in mitochondria. The autonomously replicating form of Kalilo DNA is designated AR-Kal DNA and described as the Kalilo plasmid. The inserted form of the Kalilo plasmid in the mtDNA is named mtIS-Kal DNA (Bertrand et al., 1986). The Kalilo plasmid can be transferred from senescence-prone strains to normal strains by sexual transmission or by transient hyphal anastomoses of the strains of the same mating types, but not between strains of opposite mating types (Bertrand, 1987).

Somehow the extrachromosomal *Kalilo* plasmid enters the mitochondria and elicits senescence by inserting into the mtDNA. In different strains the sites of insertion in mtDNA are different. These insertion sites also vary from subculture series to subculture series of the same strain (Bertrand *et al*., 1985; Bertrand *et al*., 1986). Another characteristic feature of the mtIS-Kal DNA is the fact that all of these are flanked by large inverted repeats of mtDNA (Bertrand, 1987; H. Bertrand, pers. comm.). This insertion of a foreign Kal DNA sequence in mtDNA causes mutations in the mitochondrial genome which, in turn, result in accumulation of "suppressive" mutant mtDNAs. Thus, certain types of mitochondrial mutations, rather than the presence of Kal DNA itself, cause the defective mtDNAs to accumulate. Therefore, it has been suggested that the role of the *Kalilo* plasmid as a senescence-inducing factor is due to its activity as a mutator of mitochondrial genes, rather than the possibility that *Kalilo* has an origin of replication that provides a replicative advantage to mtDNAs that contain mtIS-Kal DNA sequences (Bertrand *et al*., 1986; Bertrand, 1987).

Maranhar is another nuclear plasmid found in an *N. crassa* strain collected from Aarey, India (H. Bertrand, pers. comm.). This is a 6.7 Kb linear plasmid having long terminal direct repeats at its ends. The 5' ends of this plasmid, like *Kalilo*, are also protected by covalently bound protein. There is no homology between the *Maranhar* and *Kalilo* plasmids, yet *Maranhar* also inserts into the mtDNA and causes senescence. Despite the presence of direct repeats at its ends when it inserts into mtDNA, it generates long inverted repeats of the mtDNA at the point of insertion, as does *Kalilo* (H. Bertrand, pers. comm.).

7) Vectors for *Neurospora* transformation

The demonstrated ability to construct plasmids capable of replicating in yeast and bacteria, by inserting a selectable yeast marker and a yeast autonomously replicating sequence (ARS) into a bacterial plasmid such as pBR322, has prompted a number of attempts to construct similar vectors for filamentous fungi (Kinsey, 1985). At the time this thesis project was initiated there was reason to believe that the LaBelle mitochondrial plasmid, a plasmid endogenous to *Neurospora*, might be useful as part of a shuttle cloning vector for *Neurospora* species (Stohl *et al.*, 1982). Hence it was deemed important to know the complete DNA sequence and other genetic characteristics, such as the possible presence of genes which might be useful as selectable markers, in order to construct an optimum cloning vector. While this project was underway Stohl and Lambowitz (1983) utilized the *qa* system, the LaBelle plasmid, and pBR325 to construct a shuttle vector for *Neurospora*. This construct called pALS1, was used to transform a *qa*-2 *arom*-9 double mutant strain. The initial results indicated that the hybrid plasmids containing LaBelle plasmid sequences were more efficient at transformation than were equivalent plasmids lacking the LaBelle sequences. They were able to show that pALS1 replicates autonomously in both *Neurospora* and *Escherichia coli*. In addition to pALS1 they recovered a smaller deletion derivative of pALS1 called pALS2. Restriction analysis of the pALS2 plasmid indicated that most or all of the LaBelle plasmid sequences had been deleted. In fact, subsequently Stohl *et al*., (1984) showed that pALS2 that was repeatedly recovered from *Neurospora* transformants, completely lacked LaBelle sequences. These results suggested that pBR325 plus the *qa-2*$^+$ segment constitutes a *Neurospora* replicon. Similar results were obtained by Kuiper and de Vries in 1985. They constructed a vector, namely pMK2, consisting of the LaBelle mitochondrial plasmid, a part of the *qa* gene cluster of *Neurospora* and the *E. coli* plasmid pBR322. Compared to plasmid pVK88, which did not contain the LaBelle sequence, the pMK2 plasmid gave a five-fold increase in transformation of the *qa-2*$^+$ gene. In most cases the *qa* insert of pMK2 was integrated into the nuclear DNA of the host and, like pALS1, this plasmid pMK2 was also unstable. The *qa* insert as well as the LaBelle sequence of pMK2 were rapidly lost from the plasmid. They frequently recovered plasmid pMK2-5, which was a dimer of pBR322, carrying two 950-bp PstI fragments of the mtDNA EcoRI-4 fragment joining the two pBR322 segments at the PstI sites. These PstI fragments correspond to the first 550 bp of the *oxi-2* gene and an additional 400 bp upstream of the gene. Thus again, pBR322 plus the *oxi-2* PstI fragment from mtDNA EcoRI-4 constitutes a replicon in *Neurospora*. Hence, it was clear that the LaBelle mitochondrial plasmid may not be the most useful part of a shuttle cloning vector for *Neurospora* transformation, and, in further efforts to develop such a vector, the LaBelle mitochondrial plasmid was not utilized.

Currently the most popular vector used for *Neurospora* transformations is pSV50. This cosmid vector can harbour large

Neurospora inserts and contains the dominant selectable gene of N. crassa which confers benomyl resistance (Vollmer and Yanofsky, 1986). A genomic library made in this vector and arranged in order in the wells of microtiter dishes as individual constituent clones, allows the utilization of the sib-selection procedure to clone any gene corresponding to any Neurospora mutant for which an appropriate selection can be devised. In the sib-selection procedure, a gene is isolated by successive rounds of transformation of a strain of Neurospora, using pools of plasmid DNAs which are progressively reduced in complexity until a single candidate clone is identified (Akins and Lambowitz, 1985).

Over the years, DNA sequence analysis has provided an enormous wealth of information about many diverse aspects of many different organisms. The DNA sequence analysis of the Mauriceville mitochondrial plasmid had revealed much information about its genetic organization and coding potential. The DNA sequence data indicated that this plasmid was a novel genetic element in the Neurospora mitochondria. Based on this it was suggested that this plasmid might be a member of a class of genetic elements that were or are the progenitors of mtDNA introns (Nargang et al., 1984). Since, the LaBelle plasmid was shown to be distinct from any other Neurospora mitochondrial plasmid, the task of determining the complete DNA sequence of this plasmid was undertaken in the hope that this information might shed some light on its origin, biology, and functions.

CHAPTER II

MATERIALS AND METHODS

(1) Strains of *Neurospora*

Three strains of *Neurospora* were used in this work: *Neurospora intermedia* strain LaBelle-1b (P405 or FGSC #1940), referred to in this thesis as NCN47, *N. intermedia* strain Varkud-1c (FGSC #1823, *A*), referred to in this thesis as NCN50 and *Neurospora crassa* 240 *A* *nic-1* *al-2*, referred to in this thesis as NCN10.

(2) Strains of *E. coli*

Three strains of bacteria were used in this work:

E. coli strain HB101 (Maniatis *et al.*, 1982) F^-, *hsdS20* (r^-_B, m^-_B), *recA13*, *ara-14*, *proA2*, *lacY1*, *galK2*, *rpsL20*, *xyl-5*, *mtl-1*, *supE44*, λ^- was obtained from Dr. H. Bertrand, University of Regina. This strain was used as host for plasmids pLSP1 and pLSLE38 (see below).

E. coli strain JM103 (Messing, 1983; Yanisch-Perron *et al.*, 1985) Δ (*lac pro*), *thi*, *strA*, *supE*, *endA*, *sbcB*, *F'traD36*, *proAB*, *lacI*^{q}, *Z*Δ*M15* was obtained from Dr. J. Messing. This strain was used as a host for M13 transfections.

E. coli strain JM83 (Yanisch-Perron *et al.*, 1985) *r-* *ara* Δ (*lac pro*), *rpsL*, *thi*, *ϕ80*, *dlac*, *Z*Δ*M15*, *hsdR*$^-$ was kindly provided by Dr. R.A. Kelln, University of Regina. This strain was used as host for pUC-derived plasmid transformations.

(3) Recombinant plasmids

The EcoRI-6 mtDNA fragment of *N. intermedia* strain LaBelle-1b was cloned into the EcoRI site of the bacterial plasmid pUC19. This recombinant plasmid (pNK1) was used as a source of the EcoRI-6 mtDNA fragment of N. intermedia strain LaBelle-1b. The recombinant plasmid containing the EcoRI-6 fragment of mtDNA from N. crassa strain 74-OR23-A which was cloned into the EcoRI site of pBR322 to give plasmid pHBE6, was a gift from Dr. H. Bertrand, University of Regina. The recombinant plasmids pLSLE38 and pLSP1 were used to facilitate large-scale isolation of mitochondrial plasmid DNA. pLSLE38 contains the 4.1 Kb LaBelle mitochondrial plasmid cloned into the EcoRI site of pBR322, and pLSP1 contains the LaBelle plasmid cloned into the PstI site of pBR322. These two recombinant plasmids were obtained from Drs. Lori L. Stohl and Alan M. Lambowitz, St. Louis University. The HindIII-7c fragment of EcoRI-3 mtDNA which contains the cytochrome c oxidase subunit 1 gene was cloned into the HindIII site of the pUC19 plasmid. This recombinant plasmid pEL6 was generously provided by fellow graduate student Mr. Edmond G. Lemire of our laboratory.

(4) **Maintenance of the strains**

Neurospora strains were maintained by standard procedures (Davis and de Serres (1970). Working stocks were maintained on slants containing solidified Vogel's medium (see below) plus appropriate supplements. Stocks for long term storage were maintained on silica gel (6-12 mesh, grade H-Type II, dessicant, Matheson Coleman and Bell).

Bacterial colonies were maintained on appropriately supplemented agar plates which were tightly wrapped in parafilm and stored inverted at 4°C for routine usage. Bacterial strains were stored for long terms in 15% glycerol (Maniatis et al., 1982) at -20°C.

(5) Growth media

(i) Vogel's medium (Vogel, 1964)

Vogel's medium was used for vegetative growth of all Neurospora strains, both in liquid and solid cultures (Davis and de Serres, 1970).

(ii) L-broth (LB)

All bacterial strains were grown in rich L-broth (Lennox, 1955) containing 10 g Bacto-tryptone, 5 g yeast extract, 5 g NaCl, and 1 g glucose per liter. The pH was adjusted to 7.4 with 5 N NaOH. Solid medium was made by adding 15 g Bacto-agar to one liter of the above medium. For selection of ampicillin or tetracycline resistance, 100 mg/l or 20 mg/l of ampicillin or tetracycline HCl, respectively, were added to the medium after it was autoclaved and cooled to approximately 60-65°C.

(iii) Minimal medium

The JM103 strain of *E. coli* was grown in Davis minimal medium (Davis and Mingiola, 1950) without citrate, containing 10 μM thiamine and 0.5% glucose. Solid medium were made by adding 15 g Bacto-agar to one liter of the above medium.

(iv) X-Gal plates

After the autoclaved LB media (described above) was cooled to 60-65°C, 1 ml of filter sterilized IPTG (25 mg/ml dH_2O), 2 ml of X-Gal (25 mg/ml in dimethyl formamide), and 5 ml ampicillin from a 10 mg/ml stock were added. The plates were stored in the dark.

(6) Bacterial transformation

E. coli cells were made competent using $CaCl_2$ according to the method of Norgard et al., (1978). 10-20 μl of DNA solution (1-200 ng) was added to 200 μl of the competent cells. The mixture was incubated on ice for about 30 minutes, then heat shocked at 42°C for 2 minutes. After addition of 1 ml LB, the cells were incubated with shaking at 37°C for 1-2 hours. Appropriate dilutions were plated on LB containing the appropriate antibiotic and incubated at 37°C for 12-24 hours.

(7) **Large scale plasmid DNA preparation**

In the initial stages of this work, bacterial plasmid DNA was isolated from chloramphenicol-amplified *E. coli* cultures by equilibrium centrifugation in CsCl-ethidium bromide gradients, as described by Kahn et al., (1979). Using a VTi65 rotor (Beckman), the lysate was centrifuged at 54,000 rpm for 12-16 hours. The plasmid band was collected by the drip method and rebanded. Ethidium bromide was removed from the DNA by extraction with equal volumes of salt-saturated isopropanol (made by mixing 100 g NaCl; 250 ml dH_2O; 500 ml 2-propanol; 10 ml 1 M Tris-Cl, pH 7.5). This process was repeated until no ethidium bromide was apparent in the aqueous phase. The DNA solution was then dialyzed overnight against 4 l of TE (10 mM Tris-Cl, pH 7.5; 1 mM EDTA) at 4°C with one change after 2 hours. Finally, the DNA was precipitated in the presence of 0.3 M sodium acetate (pH 5.0) by 2.5 volumes of 95% ethanol. The pellet was redissolved in TE and stored frozen.

During the later stages of this work, plasmid DNA was isolated by a cleared lysate procedure according to Maniatis et al. (1982), and was further purified as per the modifications made by Akins and Lambowitz (1985). One liter cultures were inoculated with a 10 ml overnight culture and grown by shaking at 300 rpm for 4 hours at 37°C. At this time amplification was initiated by the addition of 170-200 mg chloramphenicol (dissolved in 5 ml 95% ethanol) and the cultures were incubated for another 8-20 hours. The cultures were harvested and then resuspended in 10 ml 50 mM Tris-Cl, pH 8.2; 25% sucrose; and 1 mM EDTA. A 2.5 ml lysozyme solution (10 mg/ml in 250 mM Tris-Cl, pH 8.0) was added and the mixture incubated on ice for 10 minutes. 20 ml of a solution containing 2% SDS; 50 mM Tris-Cl, pH 8.0; 62.5 mM EDTA were added and mixed by gentle inversion. This was followed by a further incubation of 15-60 minutes on ice. 2.5 ml of a 5 M NaCl solution were added to the above mixture and mixed well by inversion. The lysate was spun in a 50Ti rotor at 2°C for 45 minutes at 45,000 rpm. 0.98 gm CsCl were added for each ml of the supernatant and dissolved by gentle mixing. This solution was centrifuged at 15,000 rpm for 30 minutes at 20°C. 1 ml of ethidium bromide (10 mg/ml) was added to the supernatant and the samples were processed further using the method of Kahn _et_ _al_. (1979). After precipitation, the DNA was redissolved in 500 μl of H-NET buffer (100 mM Tris-Cl, pH 7.8; 150 mM NaCl; 1 mM EDTA), and 10 μl of RNase A (1 mg/ml, pretreated at 100°C for 10 minutes) solution were added. The solution was incubated at 37°C for 30 minutes. 50 μl of Protease K (10 mg/ml solution) were added and incubated at 37°C for 60 minutes. Water-saturated phenol was equilibrated with an equal volume of H-NET buffer. An equal volume of chloroform:isoamyl alcohol mixture (24:1 v/v) was added to this equilibrated phenol. The DNA was extracted once with this mixture followed by one extraction with chloroform:isoamylalcohol (24:1 v/v). The DNA was precipitated by the addition of 0.5 volume of 7.5 M ammonium acetate (pH 7.5) and 2 volumes of 95% ethanol. The pellet was vacuum dried and dissolved in TE.

(8) **Miniprep procedure for plasmid isolation**

Mini-plasmid preparations were done using the method of Akins and Lambowitz (1985). Cells were pelleted at 5,000 rpm for 5 minutes in an SS34 rotor from 10 ml culture. They were resuspended in 1 ml (50 mM Tris-Cl, pH 8.2; 25% sucrose; 1 mM EDTA). 250 μl of lysozyme solution (10 mg/ml in 250 mM EDTA, pH 8.0) were added and incubated on ice for 15 minutes. 2 ml of buffer containing 2% SDS; 50 mM Tris-Cl, pH 8.0; 62.5 mM EDTA were added and the mixture was incubated on ice for 15 to 60 minutes after mixing well. This was followed by the addition of 1 ml of 8 M potassium acetate and 30 minutes further incubation on ice. The lysate was centrifuged at 12,000 rpm for 10 minutes. The supernatant was transferred to clean tubes followed by addition of 1/10 volume of 3 M sodium acetate (pH 5.0) and 1 to 2 volumes of 2-propanol, depending on the volume of the supernatant. If only one volume was added, then the solution was frozen in a dry ice bath; if two volumes were added the solution was incubated for 10 minutes at room temperature. The DNA was precipitated by centrifugation in an SS34 rotor at 10,000 rpm for 10 minutes and the supernatant was removed very carefully so as not to disturb the pellet. The dried pellet was dissolved in 300 μl H-NET and 1 drop (approximately 50 μl) of RNase A (1 mg/ml) was added followed by incubation at 37°C for 30 minutes. One drop of Protease K (10 mg/ml) was then added and incubation was continued for 30 minutes at 37°C. This was followed by phenol extraction as described above (section 7) and the DNA was precipitated using 1/2 volume 8 M ammonium acetate, pH 7.5 and 2 volumes of 95%

ethanol. The dried pellet was dissolved in 200 μl TE and a second precipitation using an equal volume of 5 M ammonium acetate, pH 7.5 and 2 volumes of 95% ethanol was performed.

(9) **Transfection**

The strain JM103 was maintained on plates of Davis minimal medium + thiamine + glucose prepared as described above (section 5-iii). Overnight cultures were grown in liquid Davis minimal medium containing thiamine and glucose. 250 μl of a saturated culture were added to 25 ml (dilution of 1 in 100) of LB and grown with vigorous shaking at 37°C for approximatly 2 hours and 15 minutes so that the cells were in early to mid exponential phase. The cells were pelleted by centrifugation at 7,000 rpm in an SS34 rotor for 2 minutes and washed with 10 ml of 10 mM $CaCl_2$. The pellet was resuspended in 5 ml of 30 mM $CaCl_2$ and incubated on ice for 30 to 60 minutes. The cells were pelleted at 7,000 rpm for 30 seconds in an SS34 rotor and resuspended in 500 μl of 30 mM $CaCl_2$. Up to 20 μl (1-200 ng) of DNA solution (maximum to 1/10 volume of cells) was added to 200 μl of the competent cells thus prepared. The mixture was incubated on ice for 30 minutes and then heat shocked for 2 minutes at 42°C. Appropriate aliquots of this mixture were plated on LB plates using a 3 ml soft agar overlay containing 250 μl of lawn cells (from a fresh, saturated L-broth culture of JM103), 10 μl of 100 mM IPTG (12 mg/500 μl), and 50 μl of 2.4% X-Gal solution in dimethyl formamide (24 mg/ml). The plates were incubated at 37°C for 6-8 hours to allow the plaques to form and for color to develop due to the breakdown of X-Gal.

(10) Isolation of single stranded DNA

100 μl of an E. coli JM103 overnight culture were inoculated in a 250 ml flask containing 25 ml L-broth. To this, a single fresh plaque was added using a sterile pasteur pipette. The culture was grown with vigorous shaking at 37°C for 8.5 hours. The culture was centrifuged in an SS34 rotor at 18,000 rpm for 30 minutes to pellet the cells. The supernatant was collected and 6.5 ml of 10% PEG-6000, 2.5 M NaCl solution was added to it. After mixing, the solution was left on ice from 2 hours to overnight. The phages were pelleted by centrifuging at 10,000 rpm for 10 minutes. The phage pellet was resuspended in 650 μl of phenol extraction buffer (300 mM NaCl; 100 mM Tris-Cl, pH 7.9; 1 mM EDTA) and transferred to an Eppendorf tube. 10 μl of 10% SDS and 10 μl of Protease K (10 mg/ml) were added to the suspension, which was then incubated at 37°C for 30 minutes. This was followed by two phenol extractions and one $CHCl_3$:isoamyl alcohol (24:1 v/v) extraction. The single-stranded DNA was precipitated by addition of 800 μl of ice cold 95% ethanol and then centrifuged at 14,000 g for 30 minutes. The dried, single-stranded DNA pellet was dissolved in 100 μl of 10 mM Tris-Cl, pH 7.5 or dH_2O and stored frozen.

(11) Restriction enzyme analysis

DNA restriction endonuclease digests were basically set up as suggested by Maniatis *et al*. (1982) and the specific conditions used for different restriction enzymes were as per supplier's instructions.

(12) Agarose gel electrophoresis of DNA

DNA samples were routinely analyzed on horizontal submerged 0.8% agarose gels. A solution of 0.8% agarose (Ultra pure

TM electrophoresis grade agarose from BRL) was melted in TBE (100 mM Tris-borate, pH 8.3; 2 mM EDTA) and cast in a horizontal gel electrophoresis apparatus (Tyler Research) after the addition of 1 μl of ethidium bromide (10 mg/ml) per 100 ml of gel solution. DNA samples to be analyzed were mixed with 1/10 volume of 10X loading dye [0.15% (w/v) bromophenol blue; 0.15% (w/v) xylene cyanol; and 50% (v/v) glycerol] prior to loading in the gel slots. The gel was run in TBE buffer at 25 to 100 volts at room temperature until the desired resolution was achieved. The DNA was visualized using a UV light transilluminator (Ultra Violet Products Inc., San Gabrial, CA) and photographed with a Polaroid MP4 camera using type 57 Polaroid film and a Kodak Wratten #9 gelatin filter.

(13) **DNA recovery from agarose gel**

Three different methods were utilized to recover the DNA from the band of interest in agarose gels.

(i) Recovery of DNA from LMP gels

In the initial stages of this work, DNA was purified from low melting point (LMP Ultra Pure electrophoresis grade agarose, BRL) agarose gels. The fragment of interest was excised with minimum agarose, cut into pieces, and placed in a silanized Corex tube with an equal volume of 50 mM Tris-Cl, pH 7.8; 0.5 mM EDTA. It was then heated at 65°C until the agarose was completely melted. The sample was extracted with phenol and the phenol phase was reextracted after adding an equal volume of the above buffer. Both these aqueous phases were pooled. Following another phenol extraction and a $CHCl_3$:isoamyl alcohol extraction (as described in section 7), the DNA was precipitated from this aqueous phase.

(ii) Recovery of DNA using DE81 paper

After the band of interest was resolved in the gel a slit was made immediately before and after the band of interest with a razor blade. Excess buffer was drained and strips of dry DE81 Whatman paper (Catalogue No. 3658325, 2.5 cm ion exchange paper) precut to size were inserted in the slits. After readdition of buffer the electrophoresis was continued until the band of interest disappeared into the paper in front of it. The strip of DE81 paper containing the band of interest was retrieved, mashed into a pulp, and packed into a blue eppendorf tip. The column was washed once with 1 ml of dH_2O. Then the column was washed thrice with 400 μl each of low salt buffer containing 100 mM NaCl, 10 mM Tris-Cl (pH 7.8), and 1 mM EDTA by forcing the liquid through the column. Then, 200 μl of high salt buffer (1 M NaCl; 10 mM Tris-Cl, pH 7.8; 1 mM EDTA) was added to the column to elute the DNA. The column was allowed to drip under force of gravity overnight at room temperature. The column was washed with another 200 μl of high salt buffer and the fractions were pooled. The aqueous solution was extracted with an equal volume of water-saturated sec-butanol. The cellulose fibers were then pelleted by spinning 5 minutes in a microcentrifuge. The aqueous phase was recovered and the DNA was precipitated by addition of 800 μl 95% ethanol and then centrifuged at 14,000 g for 30 minutes.

(iii) Recovery of DNA using NA-45 membrane

The DNA from the band of interest was bound to NA-45 membrane essentially as described above. The DNA was eluted from this membrane as per the manufacturer's instruction (Schleicher & Schull).

(14) Ligations

Typically a ligation reaction was done in approximately 20 μl volume containing 1 μg of DNA. The ratio of insert:vector DNA was kept at 10:1 for vectors digested with only a single restriction enzyme and 5:1 for double digested vectors. Ligations were done in the presence of one of three ligation buffers. The first contained 50 mM Tris-Cl, (pH 7.8), 10 mM $MgCl_2$, 1 mM ATP, 2 mM spermidine, and 50 μg/ml nuclease free BSA. In the second, the BSA concentration was increased by addition of 1/10 volume of 4 μg/μl BSA for blunt-ended ligations and 1 μg/μl for sticky ended ligations, also, DTT was added to a final concentration of 2 to 10 mM. The third contained 50 mM Tris-Cl (pH 7.6), 10 mM $MgCl_2$, 5% (w/v) polyethylene glycol 8000, 1 mM ATP, and 1 mM DTT. When the ligations were done using the third buffer, the ligation mixture was diluted 5 fold in TE (pH 7.2) before transformation using the method of King and Blakesley (1986).

(15) Radiolabeling of DNA

Radioactive DNA probes for northern blots were prepared as described by Feinberg and Vogelstein, (1983, 1984), except CsCl-purified DNA was used for labeling. The labeling reaction was set up by adding the following reagents in order: dH_2O (to a total volume of 50 μl), 10 μl of oligo labeling buffer (OLB, see below), 2 μl of 10 mg/ml BSA, 5 ng DNA (heated to 95-100°C for 2 minutes followed by immediate cooling in an ice bath), 5 μl (50 μCi) $\alpha^{32}P$ dCTP (3000 Ci/mmole) and 2 units of Klenow (large) fragment of DNA polymerase I. The reaction was incubated at 37°C for 12 to 16 hours and was stopped by the addition of 200 μl of stop mix containing 20 mM NaCl, 20 mM

Tris-Cl (pH 7.5), 2 mM EDTA, 0.25% SDS, and 1 μM dCTP. The OLB was made by addition of 1 ml of 1.25 M Tris-Cl (pH 8.0); 125 mM $MgCl_2$ solution to 18 μl of 2-mercaptoethanol and 5 μl each of 100 mM dATP, dGTP and dTTP. The resulting solution was then mixed with 2 M HEPES, pH 6.6 and hexadeoxyribonucleotides ($Pd[N]_6$, Pharmacia)(in TE at 90 A_{260} units/ml) in a ratio of 2:5:3. The buffer was stored at -20°C and used for up to three months. The unincorporated radioactive nucleotides were removed by spun-column chromatography as described by Maniatis _et al_. (1982). Sephadex G-50 equilibrated with TE and 0.02% NaN_3 was used to pack a 1 ml column. The column was washed with TE three times before loading the sample. The total incorporation was determined by spotting 5 μl of the sample on to two separate circles of DE81 ion exchange Whatman filter paper. One of these was washed using 50 ml of 0.3 M ammonium formate solution which removes unincorporated radioactivity. Both samples were counted in a Beckman LS7500 liquid scintillation counter after addition of 5 ml scintillation fluid (Aqueous Counting Scintillant Amersham).

(16) **Isolation of mitochondria**

Neurospora was grown in liquid cultures essentially as outlined by Bertrand and Pittenger (1969). Cultures were grown by shaking at 25°C-30°C for 16-18 hours in Vogel's medium. Baffled flasks were used to provide good aeration. The mycelia were harvested using a Buchner funnel and Whatman No. 54 or No. 4 qualitative filter paper. The mycelial pad was rinsed with cold dH_2O and kept cold. All subsequent operations were done in the cold room or on ice. The mycelia were ground to a paste using a cold mortar and pestle with acid-washed sea sand and grinding buffer (15% sucrose; 10 mM Tris-Cl, pH 7.6; 1 mM EDTA). 1.5 gm sand and 2 ml of grinding buffer were used per gm of mycelia. Another 2 ml grinding buffer was added to suspend the ground paste. The sand, nuclei and cell debris were pelleted by centrifuging at 3,000 rpm for 10 minutes. The crude mitochondrial pellet was obtained by centrifuging the supernatant at 13,000 rpm for 30 minutes in an SS34 rotor or 12,000 rpm for 1 hour in a GSA rotor.

In the initial stages of this work, crude mitochondrial preparations were further purified using step gradients using the method of Nargang and Bertrand (1978). Flotation gradients, as described by Lambowitz (1979), were used in the later stages of work.

(17) **Isolation of mitochondrial plasmid DNA**

Mitochondrial plasmid DNA was isolated essentially by the method of Collins _et_ _al_. (1981) with minor modifications. Briefly, flotation-gradient-purified mitochondria were suspended in 3 ml of 100 mM Tris-Cl (pH 7.6), 5 mM EDTA. SDS was added to a final concentration of 1%. This solution was extracted twice with water-saturated phenol followed by one extraction with $CHCl_3$:isoamyl alcohol (24:1 v/v). The aqueous phase was collected and dialyzed against 10 mM Tris-Cl (pH 7.6), 0.5 mM EDTA. 50 µl of RNase _A+T1_ (10 mg/ml) was added for each ml of the solution and the mixture was incubated at 37°C for 30 minutes. The DNA was then banded by equilibrium centrifugation in CsCl-ethidium bromide gradients (containing 0.98 gm CsCl for each ml of lysate) which separates the covalently closed circular mitochondrial plasmid DNA from the bulk of the mtDNA, most of which is largely broken or nicked during isolation. The lower band, corresponding to mitochondrial plasmid DNA, was collected. Ethidium bromide was removed by extraction with salt-saturated 2-propanol followed by dialysis to remove CsCl.

(18) Cytochrome spectra

Cytochrome spectra were obtained by a modification of the procedure described by Bertrand and Pittenger (1969). A crude or gradient-purified mitochondrial pellet was resuspended in 1.5 ml of 50 mM Tris-Cl (pH 7.7). 1.5 ml of 5% deoxycholate was added to dissolve the mitochondrial membranes. The sample was aliquoted equally into Eppendorf tubes and microcentrifuged for 10 minutes. The air oxidized, deoxycholate dissolved mitochondria were used in both the reference and sample cells to obtain a base line by scanning from 650 nm to 500 nm. The sample was reduced by adding sodium dithionite and the reference was fully oxidized by addition of potassium ferricyanide. The scan was repeated to obtain the oxidized vs. reduced difference spectrum.

(19) DNA sequencing

The DNA sequence was determined by the dideoxy method of Sanger et al. (1977). The M13 phage derivatives mp10, mp11 (Messing, 1983), mp18 and mp19 (Norrander et al., 1983) were obtained from Dr. J. Messing and were used for the production of single-stranded template DNA as described before (section 10). The single-stranded DNA template to be sequenced was hybridized to a universal sequencing primer (New England Biolabs 15mer or 17mer). The hybridization was performed by mixing 1 to 2 μg of template DNA, 5 ng primer, 1 μl of 10 x Pol buffer (70 mM Tris-Cl, pH 7.5; 100 mM $MgCl_2$; 500 mM NaCl), and adjusting the volume to 13 μl with dH_2O. These reagents were mixed well, sealed in a glass capillary tube and then immersed for 2 to 3 minutes in a boiling water bath. They were then slowly allowed to cool to room temperature. The contents of the capillary tube were transferred into a well of a microtiter plate (NUNC 5.6 cm x 8.2 cm plate). 1 μl 0.1 M DTT, 1-2 μl of $\alpha^{32}P$ dATP (NEG-012A from New England Nuclear) and 1 unit (0.5-1 μl) of the Klenow fragment were added sequentially to the same well and mixed. 3 μl of the annealed template-primer mixture was aliquoted to each of four wells labelled A, C, G, and T. 3 μl of A, C, G, T reaction mixes (see below), respectively, were added to the appropriately labelled wells and mixed thoroughly. The microtiter plate was then incubated at 37°C for 15 minutes. For each set of 4 reactions, a chase mix was prepared by combining 6 μl of 1 mM dNTP's and 1.5 units (1.5 μl) of Klenow fragment in an Eppendorf tube. 1.5 μl of this prepared chase was added to each of the 4 reactions, mixed and the microtiter plate was incubated further for 15 minutes at 37°C. Finally, 14 μl of a stop solution (made by mixing 9.5 ml deionized formamide, 10 mg xylene cyanol, 10 mg bromophenol blue and 500 μl of 250 mM Na_2EDTA, pH 8.0) was added to each reaction. The reactions were then transferred to labelled Eppendorf tubes and boiled for 2 to 3 minutes. Approximately 2 μl of each reaction was loaded into wells of a 6% DNA sequencing gel (gels described in detail in section 20).

The A, C, G, and T reaction mixes were initially purchased from New England Biolabs. Later in this study they were prepared by mixing appropriate amounts of dNTP's and ddNTP's. The dNTP's were stored as 10 mM stocks at -70°C and diluted to 0.5 mM just prior to use. The A reaction mix was made by mixing 50 μl of 0.5 mM dCTP, 50 μl of 0.5 mM dGTP, 50 μl of 0.5 mM dTTP, 0.5 μl of 10 mM ddATP, 80 μl of 10 x Pol buffer (described above) and 360 µl of dH_2O.

The C reaction mix was made by mixi[illegible] µl of 0.5 mM dCTP, 50 µl of 0.5 mM dGTP, 50 µl of 0.5 mM dTTP, 4 µl of 10 mM ddCTP, 80 µl 10 x Pol buffer and 400 µl of dH_2O.

The G reaction mix was made by mixing 50 µl of 0.5 mM dCTP, 8 µl of 0.5 mM dGTP, 50 µl of 0.5 mM dTTP, 8 µl of 10 mM ddGTP, 80 µl of 10 x Pol buffer and 400 µl of dH_2O.

The T reaction mix was made by mixing 50 µl of 0.5 mM dCTP, 50 µl of 0.5 mM dGTP, 8 µl of 0.5 mM dTTP, 16 µl of 10 mM ddTTP, 80 µl of 10 x Pol buffer and 400 µl of dH_2O.

Chase solution was made by mixing 80 µl each of 10 mM dNTP and 480 µl of dH_2O.

When it was desirable to read greater than 200 nucleotides from a single gel then "long" sequencing mixes were used to perform the reactions. These contained reduced amounts of ddNTP's compared to the regular reaction mixes and were made as above except that 0.1 µl or 0.2 µl ddATP was added to the A mix, 1 µl of ddCTP was added to the C mix, 2 µl ddGTP was added to the G mix and 3 µl of ddTTP was added to the T mix instead of the amounts used before.

Sometimes reaction mixes contained dITP instead of dGTP to resolve a region of compression (Mills and Kramer, 1979; Sanger *et al.*, 1982). These dITP reaction mixes were made as follows: The dITP A mix was made by mixing 20 µl each of 0.5 mM dCTP, dTTP and 10 x Pol buffer; 100 µl 0.5 mM dITP; 40 µl dH_2O and 0.25 µl 10 mM ddATP.

dITP C mix was made by mixing 2 µl 0.5 mM dCTP; 100 µl 0.5 mM dITP; 20 µl each of 0.5 mM dTTP and 10 x Pol buffer; 60 µl dH_2O and 1.5 µl 10 mM ddCTP.

The dITP G mix contained 20 μl each of 0.5 mM dCTP, 0.5 mM dTTP and 10 x Pol buffer; 3 μl of 0.5 mM dITP; 140 μl of dH_2O and 1.5 μl of 1 mM ddGTP.

dITP T reaction mix was made by mixing 20 μl 0.5 mM dCTP; 100 μl 0.5 mM dITP; 2 μl of 0.5 mM dTTP; 20 μl of 10 x Pol buffer; 60 μl dH_2O and 4 μl 10 mM ddTTP.

(20) DNA sequencing gels

Routinely, 6% polyacrylamide, 7 M urea gels were used for DNA sequencing. The gels were cast in glass plates using 0.25 mm spacers. Short gels were 40 cm x 20 cm and long gels were 80 cm x 20 [illegible] Gel solution containing 7 M urea, 6% acrylamide, and 1 X TBE pH [illegible].3 (section 12) was filtered through a Sartorius membrane filter Type SM P-08 pore size 0.2 μm. 250 μl of 10% ammonium persulfate and 10 μl of TEMED were added to 40 ml of gel mix (for short gels) and the mixture was quickly poured between the glass plates. The comb was inserted to make the slots and the gel was kept flat for about 2 hours prior to use. The sequencing gels were run in 1 X TBE (pH 8.3), in vertical gel units (Tyler Research) at 35 watts constant power for the desired length of time. When 80 cm x 20 cm gels were run, 0.8 X TBE (pH 9.0), was sometimes used, both in the gel and the running buffer. After a run was complete one of the glass plates was removed, the gel covered with Saran Wrap and exposed to Kodak XAR-5 film in a film holder at -20°C. The autoradiograms were developed after adequate exposure.

(21) DNA sequence analysis

The DNA sequence was analyzed using the computer programs of Staden (1977, 1978, 1979, 1980a, 1980b) and Queen and Korn (1984).

(22) Isolation of RNA from mitochondria

The mitochondria were purified using flotation gradients as described above except that polyvinyl sulphate 25 μg/ml; spermine 35 μg/ml; 5 mM DTT; 0.5% β-mercaptoethanol; 4 mM EGTA; 5 mM N-ethylmaleimide were added to all the buffers used to purify mitochondria (Spradling et al., 1975). Total mtRNA free of cytosolic RNA contamination was prepared directly from modified flotation-gradient-purified mitochondria (as above) by extraction using the sodium dodecyl sulphate (SDS)-diethylpyrocarbonate (DEPC) method of Solymosy et al., (1968) with modifications. The post gradient purified mitochondrial pellet was resuspended in 1 ml of buffer containing 0.44 M sucrose (Ultrapure), 50 mM Tris-Cl (pH 7.6), 5 mM EDTA, 25 μg/ml polyvinyl sulphate, 35 μg/ml spermine, 5 mM DTT, 0.5% β-mercaptoethanol, 4 mM EGTA, and 5 mM N-ethylmaleimide. A solution containing 50 mM Tris-Cl (pH 7.6), 5 mM EDTA, 25 μg/ml polyvinyl sulphate, 35 μg/ml spermine, 5 mM DTT, 0.5% β-mercaptoethanol, 4 mM EGTA, 5 mM N-ethylmaleimide, 5% SDS, and 4% DEPC was freshly prepared. The mitochondria were lysed by adding 2 ml of this solution to the mitochondrial suspension. The lysate was incubated at 37°C for 5 minutes and then centrifuged at 10,000 rpm for 15 minutes at room temperature. The supernatant was transferred to a tube containing 0.25 g of NaCl. The salt was dissolved by inversion and the tube cooled on ice for 10 minutes before being centrifuged at 10,000 rpm for 22 minutes at 4°C. The supernatant was transferred to a clean Corex tube. Phenol, which had been equilibrated with an equal volume of H-NET (150 mM NaCl; 1 mM EDTA; 100 mM Tris-Cl, pH 7.8) was mixed with an equal volume of $CHCl_3$: isoamyl alcohol mixture (24:1 v/v) and the supernatant was extracted twice with equal volumes of this phenol:chloroform:isoamyl alcohol mixture, and once with $CHCl_3$: isoamylalcohol. 1/10 volume of 3M sodium acetate (pH 5.0), and 3 volumes of 95% ethanol were added to the aqueous phase. This was stored in a -20°C freezer until the RNA was needed. Just prior to use the RNA was recovered by centrifugation and dissolved in Milli-Q (ultrapure H_2O, purified using a Milli-Q purification unit) DEPC-treated sterile water.

(23) **Preparation of labelled riboprobes**

The 3.5 Kb EcoRI-BamHI fragment of the LaBelle plasmid (see results) was cloned into the Bluescribe cloning vector obtained from Stratagene Cloning Systems to yield pBSA15. The 555 bp EcoRI-BamHI fragment of the LaBelle plasmid cloned into Bluescribe vector yielded the recombinant plasmid pBSA1. These recombinant plasmids were transcribed under the conditions specified by the supplier to yield labelled RNA of very high specific activity. Briefly, the recombinant plasmid to be transcribed was linearized with the appropriate restriction enzyme. The transcription reaction was carried out in a volume of 50 μl. The following reagents were added in order at room temperature. DEPC-treated Milli-Q water (to a final volume of 50 μl), 10 μl of 5 x transcription buffer containing 200 mM Tris-Cl (pH 8.0), 40 mM $MgCl_2$, 10 mM spermidine $(HCl)_3$, 125 mM NaCl; 1 μg of linearized template DNA; 2 μl each of 10 mM ATP, GTP and CTP; 2 μl of 0.5 mM UTP; 1 μl of 0.75 M DTT; 30 units RNasin (1 μl); 5 μl $\alpha^{32}P$ UTP (50 μCi, Amersham P.B. 1023) and 1 μl T3 or T7 RNA Polymerase as required (20 units, BRL). Reactions were incubated at 37°C for 1 hour. Template

DNA was digested by 1 µl RQI DNase from Promega Biotech at 37°C for 10 minutes. After phenol extraction , performed as described before (section 22), the samples were spin dialyzed on Sartorious filters of pore size 0.05 µm for 45 minutes to remove the unincorporated nucleotides. The specific activity of these probes was determined as described before (section 15).

(24) **RNA electrophoresis**

RNA samples were electrophoresed in formaldehyde-agarose gels as described by Chung _et al_., (1981) with some modifications. All operations with these gels were carried out in a fume hood. A solution of 1.5% (w/v) agarose in RNA electrophoresis buffer (20 mM EPPS, pH 8.0; 10 mM sodium acetate; 0.1 mM EDTA) was melted and cooled to 60 °C. Formaldehyde was added to a final concentration of 10% (v/v) from a 37% (w/w) stock solution. A 3 mm thick gel was cast in a horizontal submarine gel apparatus.

The RNA sample to be electrophoresed was at a concentration of 4-5 mg/ml. 4 µl of sample, containing the appropriate amount of RNA, was added to a denaturation solution of 50% (v/v) formamide, 30% (v/v) stock formaldehyde which is 37% (w/w), 20 mM EPPS (pH 8.0), 10 mM sodium acetate, and 0.1 mM EDTA to give a final volume of 40 µl. The mixture was then heated at 65°C for 2 to 5 minutes. After cooling, 4 µl of 50% (v/v) glycerol containing 0.15% (w/v) bromophenol blue and 0.15% (w/v) xylene cyanol was added to each sample. The gel was submerged in RNA electrophoresis buffer plus formaldehyde (10% v/v). The samples were loaded and electrophoresed toward the anode at 100 volts constant voltage for about 30 minutes. Once both the dyes entered the gel, the buffer was recirculated between the two reservoirs for the length of the run using a peristaltic pump to prevent an ion gradient from forming during the run. The samples were electrophoresed at 50 volts for 12 to 18 hours.

After the run was completed, the lanes containing RNA size markers were cut from the gel and stained in a fresh solution of 500 mM ammonium acetate containing 0.5 μg/ml ethidium bromide for approximately 2 hours with gentle shaking. The slices were then washed for 30 minutes in 500 mM ammonium acetate and the RNA was visualized using a long wavelength UV light transilluminator and photographed as described before (section 12).

(25) **Northern blotting**

RNA was transferred from agarose formaldehyde gels to Gene Screen Plus membrane by blotting in a high salt solution (Thomas, 1980 and 1983). The formaldehyde gel was placed on a wick of 3 MM Whatman paper saturated in 10 x SSC (1.5 M NaCl; 0.15 M sodium citrate) that extended into a reservoir of the same. The gel was overlaid with a sheet of Gene Screen Plus membrane. This membrane had been prewetted with Milli-Q water and then equilibrated in 10 x SSC. The blotting was allowed to proceed for 24 to 36 hours whereupon the membrane was removed from the gel and the positions of the rRNA standards marked on the membrane. The blot was air dried for at least 2 hours and subsequently baked for 2 hours at 80°C in a vacuum oven.

(26) **RNA-RNA hybridization**

RNA, labelled to high specific activity as described above, (section 23) was hybridized to mtRNA immobilized on Gene Screen Plus membrane exactly as recommended by Stratagene Cloning Systems, the suppliers of Bluescribe cloning vectors. Prehybridization was done for 2 hours at 65°C by shaking in the presence of 10 ml of solution containing 50% deionized formamide, 5x SSC (0.75 M sodium chloride - 0.075 M sodium citrate), 1x PE (see below), and 150 μg/ml denatured salmon sperm DNA. 5x PE is 250 mM Tris-Cl (pH 7.5), 0.5% w/v sodium pyrophosphate, 5% SDS, 1% polyvinylpyrolidone (mw 40,000), 1% Ficoll (mw 400,000), and 25 mM EDTA dissolved in 80% of the final volume with DEPC-treated water by heating to 65°C and then cooled to 37°C, 5% BSA (fraction V) was then added to a concentration of 1%. Once all the components were fully dissolved, the solution was again heated to 65°C for 15 minutes. This solution was stored at room temperature.

Hybridization was performed in the same prehybridization solution at 65°C for 48 hours. Two to three million counts of RNA per lane to be probed were added to the bag after the probe was denatured by boiling for 2-3 minutes in 400 μl of the prehybridization solution. The blots were rinsed in 2 x SSC at room temperature with shaking for 5 minutes. Washing was done twice in 2 x SSC, 0.1% SDS and twice in 0.1 x SSC, 0.1% SDS for 15 minutes each at 65°C with shaking (1 x SSC is 0.15 M NaCl - 0.015 M sodium citrate pH 7.0).

(27) **Autoradiography**

Hybridized RNA blots were exposed to Kodak XAR-5 film in an X-ray cassette (Picker) mounted with Cronex intensifying screens (Dupont) at -70°C. After sufficient exposure the film was developed manually using Kodak X-ray Developer and Kodak Rapid Fixer following the manufacturer's instructions.

CHAPTER III

RESULTS AND DISCUSSION

I Analysis of the LaBelle Plasmid

(1) DNA sequence of the LaBelle mitochondrial plasmid

The LaBelle mitochondrial plasmid was subcloned into appropriate phage M13 derivatives. Both strands of the LaBelle mitochondrial plasmid were completely sequenced. In at least one strand, the sequence was also determined through all the restriction sites used for cloning. Sanger's dideoxy chain termination method (1977) was used for determination of the entire sequence. Fig. 1 shows the sequencing strategy used. The unique EcoR1 site is arbitrarily designated as position 1 of the sequence. Fig. 2 shows the DNA sequence of the LaBelle mitochondrial plasmid in linear form starting at the unique EcoR1 site.

The plasmid contains 4,070 bp. It has a base composition of 65.1% A+T and 34.9% G+C and is thus slightly more A+T rich than Neurospora mtDNA (60% A+T) and two other Neurospora mitochondrial plasmids Mauriceville (58.6% A+T) and Varkud (58.3% A+T) as shown in Table 2. But, when compared with Neurospora nuclear DNA (46% A+T), the LaBelle mitochondrial plasmid is highly A+T rich (Table 2). Fig. 3 shows a circular restriction map of the LaBelle mitochondrial plasmid that summmarizes the main features of the structural and genetic organization (discussed below) determined by DNA sequencing.

(2) G+C rich regions in the LaBelle plasmid

Table 3 lists the G+C rich regions found in the LaBelle plasmid arranged in decreasing order. In a sequence of overall 34.9% G+C richness, there are 16 regions of 20 base pairs or more which are 60-75% G+C rich. These G+C rich clusters are distributed all over the plasmid as shown in Fig. 3. The region "m" (Table 3, Fig. 3), (nucleotides 3700 - 3722), which has the highest G+C content in the plasmid, is found 28 nucleotides away from the most stable palindrome (3612 - 3672) (discussed later in section 7) in the LaBelle plasmid.

(3) The LaBelle plasmid contains a long open reading frame

Computer translation of the DNA sequence using the Neurospora mtDNA genetic code (UGA= tryptophan, Heckman et al., 1980) showed that the LaBelle plasmid contained several ORFs. The positions of these ORFs on the plasmid are shown in Fig. 4. The longest ORF could encode a protein as long as 1,151 amino acids (131,727 daltons) and is located in the top strand of Fig. 2, beginning with the AUG codon at nucleotides 59 to 61 and ending with the UAG codon at nucleotides 3,512 to 3,514 (Figs. 3 and 4). The AUG codon is the only one known to initiate translation in Neurospora mitochondria (de Jonge and de Vries, 1983; Macino and Morelli, 1983; Browning and RajBhandary, 1982; Morelli and Macino, 1984; Burger and Werner, 1985; de Vries et al., 1986; Nelson and Macino, 1987). However, it should be noted that the long ORF is actually contiguous to an upstream inframe codon specified by nucleotides 4,012 to 4,014 and thus contains 39 additional codons upstream from the first AUG codon. Fig. 5 shows the amino acid sequence of the polypeptide that would be encoded by this ORF. This ORF alone accounts for 85% (3,456 nucleotides) of the plasmid's coding capacity and is hereafter referred to as the long ORF. The other two reading frames in the same region of DNA contain 110 and 77 stop codons. It is extremely unlikely than an ORF of this length would exist by chance alone[1].

The predicted 1,151 amino acid polypeptide has a polarity index of 49%, calculated as described by Capaldi and Vanderkooi (1972). Since a polarity index of 47± 6% is indicative of a soluble protein (Capaldi and Vanderkooi, 1972) the predicted polypeptide would be a soluble protein. This hydrophilic polypeptide has an excess of basic over acid amino acid residues (ratio 1.34).

The hydropathy plot of the 1,151 amino acid polypeptide was obtained using the method of Hopp and Woods (1981). A running average of the amino acid hydrophilicity values are plotted as shown in Fig. 6. Examination of this hydropathy plot does not show the presence of any characteristic membrane binding domains, instead it further supports the prediction that this would be a soluble hydrophilic protein. The secondary structure of this predicted polypeptide was analyzed using the algorithm of Garnier _et al_. (1978). Calculation of the protein secondary structure predicts that 42% of the amino acids may contribute to α-helical foldings (the percentage of α-helices in proteins varies from 0-75%; the average protein is approximately 30% α-helical). In most globular proteins α-helices are not longer than seven to ten residues (Schultz _et al_., 1974; Schultz and Schirmer, 1979). The deduced polypeptide has 28 stretches of at least seven or more residues capable of forming an α-helix as shown in Fig. 6. The longest stretch of α-helix is 41 residue long and extends from amino acid 477 to 517. The polypeptide also has 5 regions containing 10 or more amino acids which are capable of forming β pleated sheets as shown in Fig. 6.

---

[1] Using the _Neurospora_ mitochondrial genetic code, where only UAA and UAG are stop codons (Heckman _et al_., 1980), a sequence of 1,152 codons should randomly contain 36 stop codons. However, since the sequence is 65.1% A+T, 61 stop codons are expected in a sequence of this length.

The long ORF of the LaBelle plasmid contains 38 AUG codons. Conceivably, any of these internal AUG codons could be used as initiating codons for a polypeptide. The properties of each potential polypeptide were examined to determine if any of them differ substantially from the longest possible ORF. The overall character of the polypeptide basically remains the same. In all cases the resulting polypeptides would still be hydrophilic and basic however, it should be noted that the carboxy terminus of the polypeptide is relatively more basic in character (Table 4).

In this analysis of the potential coding capacity of the LaBelle plasmid using the mitochondrial genetic code, several ORFs smaller than the long ORF (described above) were found. Only the ORFs, which consist of 40 codons or more (including the initiation and stop codon), were considered. Several ORFs, that are much shorter in length than the longest ORF, were found to exist on the same strand as the longest ORF but in a different frame. All these ORFs have a coding capacity of 59 amino acids or less and are of doubtful significance (Fig. 4 and Table 5).

The bottom strand of the LaBelle plasmid contains seven ORFs which could encode polypeptides of 57 amino acids, 40 amino acids, 70 amino acids, 63 amino acids, 118 amino acids and 59 amino acids (Fig. 4 and Table 5). However, all but the ORF coding for the peptide composed of

57 amino acids are present in the same region of the plasmid DNA that potentially encodes the long ORF. Although there are a few known cases where genes are present on opposite strands of the same region of DNA in eukaryotes (Henikoff *et al*., 1986; Chen *et al*., 1987; Adelman *et al*., 1987), it seems highly unlikely that all these ORFs are expressed. Furthermore, the results of transcription studies (section 12) indicate that only the top strand of the LaBelle plasmid DNA (containing the long ORF) is transcribed.

If the plasmid is translated using the standard genetic code, the long ORF would be divided into five shorter ORFs with the capacity to code for proteins of 549 amino acids, 29 amino acids, 82 amino acids, 194 amino acids and 73 amino acids (Fig. 4 and Table 5). Out of these shorter ORFs, the ORF that can encode the 549 amino acid polypeptide and possibly the ORF encoding 194 amino acids might be significant. The other ORFs, owing to their limited coding capacity, are of doubtful significance.

(4) Codon usage in the long open reading frame

In this section, exons of the mitochondrial genes which are involved in oxidative phosphorylation will be referred to as standard mitochondrial genes. Any ORFs that are present in the introns of these standard mitochondrial genes, and for which no specific function has been assigned so far, will be referred to as intronic ORFs.

Table 6 summarizes the frequencies of codon usage in the long ORF. There are two noticeable features. First, the plasmid ORF shows a bias similar to that of *Neurospora crassa* mtDNA coding sequences (both introns and exons) against codons ending in C or G. As shown in

Table 7, codons ending in C or G comprise only about 29% of the total in the plasmid long ORF, and about 17% in Neurospora mtDNA genes. Second, as shown in Table 8, the usage of certain distinctive codons in the plasmid is remarkably similar to that characteristic of ORFs in fungal mtDNA introns. The latter use CGN (arg), UGG (trp), ACC and ACG (thr), CUC (leu), and AAG (lys) codons, whereas these codons are used infrequently or not at all in standard mitochondrial genes (Hudspeth *et al*., 1982; Burke and RajBhandary, 1982; Burke *et al*., 1984; Nargang *et al*., 1984; de Vries *et al*., 1986; Nelson and Macino, 1987). On the other hand, UUU and UUC (phe codons) are in a ratio of 2:1; UAU and UAC (tyr codons) occur in a ratio of 3:1; and CAU and CAC (his codons) are present in a ratio of 4:1 in this ORF. This distribution of phe, tyr and his codons is identical to that seen in the standard mitochondrial genes, whereas in the intron ORFs they are more highly biased toward UUU, UAU and CAU codons respectively. Thus, the LaBelle plasmid long ORF resembles both intron ORFs and mtDNA exons in certain respects.

Fungal mitochondrial introns are thought to have originated by invasion of the mitochondria by foreign genetic elements and their subsequent insertion in the mtDNA (Burke *et al*., 1984; Dujon *et al*., 1987). The Mauriceville mitochondrial plasmid has been suggested to be related to these purported intron progenitors (Nargang *et al*., 1984). As described above, the LaBelle plasmid seems to be related not only to the mitochondrial intronic ORFs, but also to the mtDNA exons. So it is possible that the LaBelle plasmid is related, but not identical to, these mobile genetic elements that have given rise to present day fungal mtDNA introns.

Codon usage in the LaBelle plasmid differs strongly from that in the *Neurospora* nuclear genes whose sequences have been published (*trp*-1, Schectman and Yanofsky, 1983; histones H3 and H4, Woudt *et al*., 1983; *qa*-4, Rutledge, 1984; *qa* repressor, Huiet and Giles, 1986; plasma membrane ATPase, Hager *et al*., 1986; *am*, Kinnaird and Fincham, 1983; ADP/ATP carrier, Arends and Sebald, 1984; the preprotein of the proteolipid subunit of ATPase, Viebrock *et al*., 1982; *his*-3, Legerton and Yanofsky, 1985; and *pyr*-4, Newbury *et al*., 1986). As a group, all these nuclear genes show a strong bias toward codons ending in G or C as shown in Table 7 (70% compared to 29% in the LaBelle plasmid). Furthermore, the LaBelle plasmid uses several codons (Leu, UUA; Ile, AUA; Cys, UGU) that are used infrequently or not at all in the above nuclear genes. This, strongly indicates that the LaBelle plasmid is not related to *Neurospora* nuclear genes.

A computer search was done for the consensus sequence that is characteristic of intron-encoded maturase polypeptides (Waring *et al*., 1982; Hensgens *et al*., 1983). The amino acid sequence encoded by the long ORF revealed no close homologs to the consensus sequence.

(5) **Absence of conserved sequence elements of mitochondrial introns**

On the basis of conserved sequence elements and secondary structure the fungal mtDNA introns have been divided into group I and group II introns (Michel *et al*., 1982; Michel and Dujon, 1983; Burke *et al*., 1987). Since the codon usage in the LaBelle plasmid long ORF most closely resembles codon usage in fungal mtDNA intronic ORFs, a computer search was done for identification of these consensus sequence elements of group I and group II mtDNA intronic elements in the LaBelle plasmid. This search revealed no significant homology to either group.

(6) Palindromic sequences

All the known mitochondrial origin of replication regions share one feature in common and that is a potential to form a secondary structure (Zamaroczy *et al.*, 1984). Often, inverted repeats are also associated with genetic processes such as transposition. So the LaBelle DNA sequence was examined for the presence of palindromic sequences using the following criteria: 1) a minimum of 7 correct matches in the stem with a minimum percentage of correct matches in the stem of 80% or above; 2) a maximum insertion or deletion (i.e. loop-out) of only 5 bases present in the stem; and 3) the maximum length of the central loop of the hairpin of no greater than 7 bases. The free energies of these hairpin loop structures were calculated according to the rules of Tinoco *et al.*, 1973 by the computer program of Queen and Korn (1984). Only the structures with a ΔG of -10 kcal/mol or less were considered. Fig. 7 shows the five regions of palindromic sequences found in the LaBelle plasmid having free energies of formation less than or equal to -10 kcal/mol arranged in descending order of stability.

To detect palindromic sequences that are separated by longer distances and which may not be thermodynamically as stable but long enough to play a role in genetic processes such as recombination, the maximum length of the central loop was increased to the length of the plasmid itself (4,070). This relaxed condition allowed the detection of more than 500 palindromic regions. Therefore, the stringency of the stem was increased by requiring 90% correct matches in a stem of minimum match length 12 bp and also by allowing a maximum insertion or deletion of only 3 bases in the stem. Using these criteria an additional 11 palindromic regions were detected. All these regions are listed in Table 9. Examination of this table indicates that the palindromic sequence between nucleotides 3612-3672 which is found outside of the long ORF is thermodynamically the most stable (ΔG= -29.4 Kcal/mol) and hence, most likely to form a secondary structure _in vivo_. It is worth noting that the only stretch that is 75% G+C in content in the entire plasmid (section 2) is found at a distance of 28 nucleotides from the most stable palindromic region. Further away from this G+C rich stretch is a region containing a string of 7A's (nucleotide position 3751 to 3757). Some of these features are reminiscent of transcription termination sites in several prokaryotes (Rosenberg and Court, 1979). In plastid genomes of eukaryotes it has been suggested that inverted repeats function as transcriptional terminators (Deno _et al_., 1984; Sugita and Sugiura, 1984). On the other hand, now it has been shown that the inverted repeat sequences found at the 3' ends of the plastid DNA transcription units are ineffective as transcription terminators, but serve as efficient RNA processing elements (Stern and Gruissem, 1987). Similarly, in diverse organisms such as photosynthetic bacteria, nonphotosynthetic bacteria and animal cells it has been suggested that hairpin loop structures function as RNA processing sites or protective structures against nucleolytic degradation (Wong and Chang, 1986; Newbury _et al_., 1987; Belasco _et al_., 1985; Mott _et al_., 1985; Birchmeier _et al_., 1984). In view of the location of this palindromic sequence outside of the longest ORF and the fact that the long ORF is transcribed (discussed in section 11 and 12) the most attractive hypothesis for this region of the LaBelle mitochondrial plasmid might be as an mRNA processing and stabilizing element. If this palindromic sequence does play a role in increasing the stability and subsequent processing of the transcript from the long ORF, then it might be possible that at least some of the smaller transcripts (discussed in section 12) could actually be the processed or degradation products of some of the longer transcripts. Alternatively, this palindromic sequence may also be involved in the replication of the plasmid (see conclusions).

PstI-palindromes, which are repetitive sequence elements found flanking genes in *Neurospora* mtDNA and in the Mauriceville and the Varkud plasmids (Yin *et al.*, 1981; Nargang *et al.*, 1983; Nargang *et al.*, 1984; Akins *et al.*, Submitted), were not found in the LaBelle mitochondrial plasmid.

(7) **Direct repeats**

The LaBelle mitochondrial plasmid sequence was scanned for regions containing direct repeats using the following criteria. The minimum number of correct matches required in a repeat was set at 15, with a minimum of 80% correct matches out of the total length of the repeat. An insertion or deletion (loopout) of a maximum of 3 bases was allowed in one member of the repeated pair with respect to the other. The maximum distance between the repeat units was set equal to the length of the plasmid itself (4,070) so that all possible pairs of direct repeats could be found. When no loopouts (*i.e.* insertion or deletion) were allowed, nine pairs of direct repeats were found. On allowing for insertion or deletion of one base, 23 additional direct repeats were found. Relaxing the conditions to allow for insertion or deletion of two bases allowed the detection of 15 more pairs of direct repeats. When a three base insertion or deletion was allowed, only one additional direct repeat was detected. Table 10 shows the direct repeats found in the LaBelle plasmid containing 17 or more nucleotides, with a percentage of identity 76% or greater.

One pair of direct repeats found in the LaBelle plasmid is shown in Fig. 8a. In this pair of direct repeats 82% of the bases are correctly matched between the region spanning from 3572 to 3592 and 3906 to 3927. This pair of direct repeats is found in the region flanking the most thermodynamically stable palindromic sequence (-29.4 kcal/mol) present in the LaBelle plasmid (palindromic sequence found between 3612 to 3672 nucleotide position) as shown in Fig. 3.

A noticeable feature of the direct repeats is their distribution in the LaBelle plasmid. Fig. 3 shows the distribution of the five most stable palindromic sequences in the LaBelle plasmid. One unit of the pair, from each of four different direct repeats (repeats I, II, III, IV in Fig. 3) also form part of four of these palindromic sequences. The direct repeat that is part of the most stable palindrome in the LaBelle plasmid and the pair that flank this palindrome may be important in replication (see conclusions). The significance of the distribution of the other palindromes and direct repeats is not clear. Furthermore, the 25 bp region from nucleotide 3907 to 3931 (part of the direct repeat shown in Fig. 8a), in the LaBelle plasmid shows 92% of the bases correctly matched to a part of _Xenopus laevis_ repetitive 1723 element clone C17. As shown in Fig. 8b the part of the _X. laevis_ repetitive 1723 element clone C17 that matches the LaBelle direct repeat spans from nucleotide 542 to 566 of clone C17. The 1723 element belongs to a transcribed family of sequences whose members are 6-10 Kb long, highly homologous to each other, dispersed in the genome, and flanked by short sequence duplications. The X. laevis genome contains about 8500 copies of this element (Kay and Dawid, 1983). Since it is a very small part of this 1723 element that shows this similarity to the LaBelle direct repeat it is difficult to draw any meaningful conclusion. Nevertheless, it is interesting to note that the sea urchin Csp2109B genomic, cloned, middle repetitive element contains a 25 nucleotide long sequence that is 100% similar to the region of the X. laevis 1723 that shows similarity to the LaBelle sequence as shown in Fig. 8b. This Csp2109 repetitive sequence is known to be represented in sea urchin egg poly (A) RNA, and is also transcribed to various extents into nuclear RNA, in a stage and tissue-specific manner, but is not translated (Posakony et al., 1981).

(8) Comparison of the LaBelle plasmid DNA sequence to the DNA sequences in the Gene Bank database

A search was done of the Gene Bank database to discover if any homology exists between the DNA sequence of the LaBelle mitochondrial plasmid and any other sequence. This search identified 16 regions that showed similarity to DNA sequences from different organisms. These regions of similarity had lengths ranging from 25 to 72 bp with percentage identities varying from 67.8% to 92%. Out of these 16 regions, those that might be biologically meaningful or have some feature such as association with repetitive elements, regulatory signals, or those that could be involved in a recombination process are discussed below.

The relatedness of a repeat unit to X. laevis and sea urchin repetitive sequences has been noted above (section 7). It was found that the region of LaBelle between nucleotides 3863 to 3897 were 80% identical to a region found on one side of an SV40 insertion site in the rat genome (Hasson et al., 1984). In SV40 both integrative and nonintegrative recombination has been shown to take place near short direct repeats in the virus genome (Hasson et al., 1984). In this context it is worthy to note that in these regions both rat and the LaBelle plasmid contain the repeated unit "TTG" (Fig. 8c). Whether this short repeated sequence is also involved in recombination in the LaBelle plasmid is unknown. In Aplysia, a marine gastropod, the mRNA for a neuropeptide contains sequences complementary to the same TTG repeated sequences (Fig. 8c) at its 3' untranslated region (Nambu et al., 1983). These sequences are found in the coding strand of the neuropeptide gene and show similarity to LaBelle (69.0% to nucleotides 3881 to 3938).

The pSpec 1 and pSpec 2 cDNA clones in sea urchin code for 2 related mRNAs. These mRNAs are tissue specific and developmentally regulated. They belong to a small family of mRNAs that code for approximately 10 low molecular weight acidic proteins actively synthesized in the embryonic ectoderm. The 3' untranslated region of both the Spec1 and Spec2 mRNAs contains highly repeated elements that are approximately 150 to 200 bp in length. It has been suggested that the insertion of one such element between the coding sequence of the spec mRNA ancestral gene and the transcription termination site followed by subsequent amplification of this gene gave rise to the functional family of spec related mRNAs. These repeated elements are present at an additional 2000 to 3000 sites in the genome. Some of the additional sites may be transcribed but these do not represent ectodermal proteins (Carpenter _et al_., 1982). The 47 bp region between nucleotide 3890 to 3936 of the LaBelle plasmid shows 73% identity to part of this 3' untranslated part of Spec 1 or Spec 2 mRNA. Interestingly, this region of LaBelle is outside[1] of the long ORF of the plasmid.

In _Triticum aestivum_ the α, β and γ gliadins are alcohol-soluble storage proteins of wheat endosperm. These gliadin proteins are unusual in that they contain 2 domains that are rich in polyglutamine (up to 20 to 30 amino acid long) coded by CAA codons (Okita _et al_., 1985; Sumner-Smith _et al_., 1985). The 20 nucleotides in the bottom strand from position 3944 to 3863 of the LaBelle plasmid show similarity (67.8% to 71.2% identity) to the regions of the α, β and γ gliadin protein encoding sequences containing the polyglutamine stretches.

At least 6 out of 17 actin genes present in the _Dictyostelium discoideum_ genome are transcribed (McKeown and Firtel, 1981). The actin 8 gene of this slime mold contains 55 nucleotides at its 3' non-coding and untranscribed region that shows 70% identity to the 56 nucleotides at positions 3553 to 3608 of the LaBelle plasmid. This region in D. discoideum is A+T rich (McKeown and Firtel, 1981) and seems to have no other special features. The region of similarity in the LaBelle plasmid is also found outside of the longest ORF in the plasmid.

---

[1] Since the LaBelle mitochondrial plasmid is circular and the 1,151 ORF accounts for 85% of the coding capacity of the plasmid, it is very difficult to decide if the sequences found in the remaining 614 bp (15%) of the plasmid are upstream or downstream of the ORF. Hence, throughout this work these sequences have been referred to as the sequences outside of the longest ORF.

The 32 bp region between nucleotides 3810 to 3841 in the LaBelle plasmid shows 81.3% identity to a part of the Vaccinia virus genome. The region of similarity corresponds to the 5' end of an ORF located immediately downstream from the DNA polymerase gene of Vaccinia virus but transcribed in the opposite direction compared to the DNA polymerase gene (Earl et al., 1986).

In Torulopsis glabrata a conserved dodecanucleotide sequence: 5'-TATAATATTCTT- 3' has been shown to be a transcript processing signal in the mitochondria of this yeast. Structures departing slightly from the consensus sequence can be active; however, the single cytosine residue in the sequence is always conserved. Motifs of the dodecanucleotide sequence where this C is not present are not active in signalling the transcript processing (Clark-Walker et al., 1985). A motif similar to the one found in this yeast is seen in the LaBelle plasmid outside of the longest ORF from nucleotide position 3574 to 3564 (as shown in Fig. 8d). If this motif were to serve as a transcript processing signal then it should be found on the transcript itself; however, in the case of the LaBelle plasmid the motif is found on the nontranscribed strand of DNA (see section 3).

Finally it should be pointed out that in Drosophila melanogaster a family of transcribed repeats called "opa" is shared among several developmently regulated loci (Wharton *et al*., 1985). Furthermore, it has been shown that sequences homologous to *opa* are found both in protein coding regions and in 5' or 3' untranslated regions of mRNAs derived from a wide spectrum of organisms, including eukaryotic viruses. Proteins encoded by these mRNAs possess diverse functions (Wharton *et al*., 1985). The *opa* repeats consist predominantly of the triplets CAG and CAA, both of which code for Gln. However, it has also been shown that regions of DNA that are homologous to *opa* do not always code for the same amino acid - *i.e.* they appear to be translated in different frames (Wharton *et al*., 1985). In this context it is interesting to note that the region of the LaBelle plasmid between nucleotides 3878 and 3932 contains several triplets that are repeated. Depending on the frame in which they are translated, they could represent repeating Thr (ACT, ACA), Tyr (TAC), Cys (TGT), Val (GTT or GTA) or Leu (TTG) residues. This region is outside of the longest ORF and partially overlaps some of the regions which were discussed above. Wildeman *et al*., (1986) have shown that the "CAT" family of repetitive DNA sequences is widely dispersed in *S. cerevisiae* , *E. coli*, *Drosophila*, man and the *N. crassa* genome. It has been suggested that both the *opa* and "CAT" repetitive elements may be involved in gene expression at the transcriptional or translational levels via a cis-acting effect (Wharton *et al*., 1985; Wildeman *et al*., 1986). If the above observations are considered together, it is not unreasonable to propose that the region found outside of the long ORF may be involved in the regulation of the expression of the putative protein.

(9) Comparison of the deduced amino acid sequences coded by the reading frames found in the LaBelle plasmid with protein sequences in the Gene Bank database

In an attempt to elucidate the function, if any, of the LaBelle plasmid, the amino acid sequences of the deduced proteins specified by the entire LaBelle plasmid (disregarding the stop codons) in all possible reading frames, were compared to the protein sequences in the Gene bank database. The Neurospora mitochondrial genetic code (UGA = Trp) was used for theoretical translation of the DNA sequence of the LaBelle plasmid.

The protein encoded by one of the reading frames in the bottom strand of the LaBelle plasmid had regions of similarity to parts of Aspergillus nidulans URFc protein, URF2 protein encoded by X. laevis mtDNA, the β chain precursor of the acetylcholine receptor protein from Torpedo californica, and the protein encoded by ORF 430 in a filamentous virus, PF3, of the gram negative bacterium Pseudomonas aeruginosa (Netzker et al., 1982; Roe et al., 1985; Noda et al., 1983; Luiten et al., 1985). In the second possible reading frame, the amino acids encoded by the bottom strand of the LaBelle plasmid showed some similarity to a region found in the middle of the coding sequence for apocytochrome b in the mtDNA of maize (Dawson et al., 1984). The protein specified by the third possible reading frame in the bottom strand revealed some identity to the SGS-3 protein of Drosophila (Garfinkel et al., 1983). These regions of identity are either short (26 amino acids) with a high percentage of perfect matches up to 57.7% or if they are longer (upto 157 amino acids) then the percentage of perfect matches is as low as 32.5%. Only 3 out of the 6 regions of similarity are found associated with ORFs of over 40 codons or more (as defined in section 3) on the bottom strand. Further, as discussed before, it is unlikely that these ORFs are expressed simultaneously along with the long ORF on the opposite strand of the plasmid. Hence it seems unlikely that any of these regions of similarity are of any significance.

When the top strand (Fig. 2) of the LaBelle plasmid was translated in all possible reading frames and compared to the database, regions of similarity were found in only one frame. Furthermore, all these regions of similarity were found to be part of the long ORF as shown in Figs. 4 and 5. Initially the only region of homology found in the long ORF was to the β' subunit of the DNA-directed RNA polymerase of E. coli. As described earlier the first met residue of the long ORF specified by the nucleotides at positions 59 to 61 is designated amino acid number 1 of the ORF (as shown in Fig. 5). Computer analysis detected a 59 amino acid region in the LaBelle long ORF, between amino acid residues 50 to 108, that showed 36% identity to a region of the E. coli RNA polymerase β' subunit (region d, Fig. 9 and Table 11). The β' subunit of the E. coli RNA polymerase is encoded by the rpoC gene and is a 1407 amino acid long polypeptide which contains the DNA binding domain of the E. coli RNA polymerase holoenzyme (Ovchinnikov et al., 1982). Closer analysis was done by dot matrix analysis followed by protein alignment by similarity using the algorithm developed by Dayhoff and colleagues (Queen and Korn, 1984). In this algorithm each pair of amino acids, based on their chemical structure and properties, is assigned a similarity score. The two proteins are aligned, with gaps inserted as needed, to maximize the total similarity of their amino acids. Thus the method produces an optimal alignment of two proteins, with non-identical but similar amino acids counted as a partial match. The results of such an analysis showed that amino acid residue numbers 3 to 119, 195 to 236, 309 to 402, 535 to 582 and 854 to 898 of the LaBelle long ORF show similarities to regions of the β' subunit of RNA polymerase of E. coli (Fig. 9 and Table 11). The percentage identity in these regions varies from 16.7% to 36.1%. At least some of these regions of similarity seen between the LaBelle long ORF and the E. coli RNA polymerase β' subunit also show the correct 5' to 3' alignment (regions a, d and f in Fig. 9). These observations suggested that the LaBelle long ORF may have some relationship to E. coli RNA polymerase. In view of the above observation a detailed comparison was performed with amino acid sequences of the other subunits of the E. coli RNA polymerase. These searches showed that there was no similarity between the LaBelle long ORF and any other E. coli RNA polymerase subunit (other than those noted above to the β' subunit).

Shinozaki et al., (1986) have shown that the Nicotiana tabacum var. Bright Yellow 4 chloroplast DNA contains ORFs and RFs (RF, a region from a stop codon to the next stop codon) which may represent a split gene version of the RNA polymerase β' subunit (rpoC) encoded by the chloroplast genome. A comparison of the residues specified by the LaBelle long ORF with these regions of the chloroplast genome revealed that amino acid residues 341 to 384 show 30.4% identity and amino acid residues 943 to 989 show 23.4% identity (regions j and r in Fig. 9 and Table 11) to two different parts of RF862 of the tobacco chloroplast genome. RF862 is part of the region in this chloroplast genome that may encode the β' equivalent subunit of RNA polymerase. (Shinozaki et al., 1986).

The sequence of the yeast mtRNA polymerase was kindly provided prior to publication by Dr. D. Clayton, Stanford University. A comparison of the amino acid sequence of the LaBelle long ORF to this sequence revealed that the amino acid residues 18 to 69, 254 to 300, 342 to 388, 418 to 461, 794 to 836, 857 to 904 and 1004 to 1067 (Fig. 9 and Table 11) specified by the long ORF show similarities that range from 20% to 34.8% identity with the yeast mtRNA polymerase core enzyme. Furthermore, some of these regions of identity were also found in the correct 5' to 3' direction (regions b, g, m, o, q and s; Fig. 9). This enzyme is encoded by the RPO41 gene on chromosome 6 of yeast. Contrary to earlier observations, Schinkel et al., (1986) were able to resolve yeast mtRNA polymerase into two different components by chromatography on Blue Sepharose. Currently it is believed that yeast mtRNA polymerase interacts with at least one other protein factor (Masters et al., 1987).

Yeast mtRNA polymerase has been shown to be homologous to the bacteriophage T3 and T7 RNA polymerases (Masters et al., 1987) which are single polypeptide enzymes (Moffatt et al., 1984 and McGraw et al., 1985). In view of these observations, a comparison was made of the amino acid sequence specified by the LaBelle long ORF with these bacteriophage RNA polymerases. Amino acid residues from 54 to 72, 113 to 149, and 572 to 611 specified by the long ORF were found to show similarity to parts of the T7 RNA polymerase (28.6% to 35.0% as shown in Table 11). However, the three regions of T7 RNA polymerase that show similarity to the protein specified by the long ORF are not parts of those regions which are conserved between the T7 and yeast mitochondrial RNA polymerases (Masters _et al_., 1987).

The T3 and T7 RNA polymerases are 82% identical in their amino acid sequence (McGraw _et al_., 1985). As shown in Table 11, three regions of the putative protein specified by the LaBelle long ORF from amino acid 469 to 502, 573 to 596, and 1078 to 1121 show 25.9% to 28.9% identity to three parts of T3 RNA polymerase. However, two of these three regions, as in the previous situation, are not the same as the regions which are conserved between the yeast mtRNA polymerase and T3 RNA polymerase (Masters _et al_., 1987). Only part of the third remaining region of T3, which is similar to (amino acid 573 to 596) of the protein specified by the long ORF of LaBelle, is conserved between yeast mtRNA polymerase and the T3 RNA polymerase.

Amino acid residues 486 to 527 of the long ORF show 22.9% identity to a region of the SP6 RNA polymerase (Table 11) which is distantly related to the T7 RNA polymerase (Kotani _et al_., 1987) at the amino acid level. The observation, that the 1,151 amino acid polypeptide coded by the long ORF has nine regions (regions a, b, d, f, g, m, o, q and s; Fig. 9 and Table 11) that show varying degrees of identity (21.6% to 36.1%) to either the β' subunit of the multisubunit _E. coli_ RNA polymerase or to the single polypeptide yeast mtRNA polymerase, raised the question of how much similarity there was between the β' subunit of _E. coli_ RNA polymerase and the yeast mtRNA polymerase. Comparison of these two sequences by dot matrix analysis followed by sequence alignments as explained above revealed the four regions shown in the bottom part of Table 11 which have percentage identity varying from 22.6% to 30.6%. Of these four, two (regions v and w; Fig. 9 and

Table 11) have the correct 5' to 3' order.

Thus, from these analyses it becomes clear that different parts of the 1,151 amino acid polypeptide encoded by the long ORF show varying degrees of similarities to various RNA polymerases. Remarkably, if this 1,151 amino acid protein specified by the long ORF was (or is) a plasmid encoded RNA polymerase then this RNA polymerase represents an unusual situation by showing similarity to both a component of a multi-subunit RNA polymerase and also to single polypeptide RNA polymerases.

If the 1,151 amino acid long polypeptide was or is a plasmid-encoded RNA polymerase then it should have a DNA binding domain to enable it to fulfill its function. Based on crystallographic data which is available only for three DNA binding proteins; λ Cro, λ repressor and catabolite gene activator protein (CAP), it has been proposed that in complexes of these proteins with DNA, many of the DNA contacts are made by two α-helices that are linked by a tight turn (Pabo and Sauer, 1984). Numerous DNA-binding proteins such as Galactose repressor, H-inversion protein, transposon Tn3 resolvase etc. have been shown to share sequence homologies with these three proteins, and have been predicted by several groups to also use helix-turn-helix structures for DNA interactions (Pabo and Sauer, 1984; Laughon and Scott, 1984 and references therein). So a search was done for the consensus amino acid sequence found in the DNA binding domain. As shown in Fig. 10 amino acids 219 to 238 of the 1,151 amino acid ORF show similarity to the DNA binding domain. In the numbering scheme used in Fig. 10 amino acid residues at positions 4, 5 and 8 in the first helix and 15, 18 and 19 in the second helix have nonpolar residues and face the hydrophobic core of the protein. In addition, residue 10 is also nonpolar in the DNA binding domain (Pabo and Sauer, 1984). The LaBelle specified polypeptide contains hydrophobic amino acids at all these positions. Position 10 contains Tyr which is expected to be an uncharged polar amino acid but in reality is the least soluble of all the amino acids in aqueous solution. The amino acid residues at positions 5, 9 and 15 are among the most highly conserved in the DNA binding domains of the proteins examined. Alanine is favoured at position 5, glycine predominates at position 9, and either valine or isoleucine usually occupy position 15 (Pabo and Sauer, 1984). The putative LaBelle ORF DNA binding domain has alanine at position 5, valine at position 15 but has isoleucine instead of glycine at position 9. Originally, it was thought that glycine was required at this position for formation of a proper turn between helices 2 and 3. But a double mutant of the λ repressor where Asp is changed to Asn at position 6 and Gly is changed to glutamic acid at position 9 is functional (Pabo and Sauer, 1984). So, it has been suggested that residues other than glycine might be accommodated at position 9 but at the cost of some modest conformational strain. Several proteins carry other amino acids such as Ser, Cys, Asn instead of Gly at this position in their DNA binding domains (Pabo and Sauer, 1984; Laughon and Scott, 1984). Therefore, even though Ile may not allow the same conformation as Gly, it might still be functional. Thus it might be that amino acids 219 to 238 of the protein specified by the long ORF comprise a DNA binding domain in this putative polypeptide.

There are limitations to this correlation. It should be noted that the crystal structure of the large proteolytic fragment of _E. coli_

DNA polymerase I complexed with dTMP shows a large cleft whose size and shape have been shown to suggest that it can bind double-stranded B-DNA (Ollis _et al_., 1985a). In this case it has been shown that despite the structural and functional analogy between the DNA binding region of the Klenow fragment and the two-helix DNA binding motifs noted above, there does not seem to be a strong structural homology (Ollis _et al_., 1985a). Furthermore, all the proteins that are shown to contain the double helical DNA binding domain described in the preceding paragraph are either regulatory proteins (like CAP, λ Rep, λ Cro) or sequence specific DNA binding proteins (like MATa1, MATα2) or homeotic gene products (like Ftz, Antp, Ubx) which are involved in _Drosophila_ development (Pabo and Sauer, 1984; Laughon and Scott, 1984).

Ollis _et al_., (1985b) have shown that common structural features are shared between the parts of the Klenow fragment from _E. coli_ and T7 DNA polymerase that bind the DNA substrate. Comparison of the amino acid sequences of the Klenow fragment of DNA polymerase I and T7 DNA polymerase with the sequence of T7 RNA polymerase lead to the identification of a few short regions of homology (6-15 residue long containing between 3 and 7 identical residues) (Ollis _et al_., 1985b). Due to this limited homology the evolutionary relationship between the DNA and RNA polymerases could not be evaluated (Ollis _et al_., 1985b). So the possibility that the various RNA polymerases have DNA binding domains that are different from those of the DNA polymerase I large fragment should be borne in mind. Thus considering the lack of extensive homology to other RNA and DNA polymerases it is possible that the LaBelle mitochondrial plasmid may encode an RNA polymerase that has diverged considerably from other known RNA polymerases or conceivably, it might encode another class of DNA binding protein.

(10) Repeated motif in the 1,151 amino acid long polypeptide

The amino acid residue motif Asp Val Val Phe Ala Phe Pro from position 306 to 312 in the 1,151 amino acid long polypeptide, is found repeated at amino acid positon 665 to 671 with one change. The motif at this position is Asp Val Val Phe Lys Phe Pro. In this seven amino acid motif only one residue Ala (nonpolar) is changed to Lys (positively charged) in the second repeated region. The significance of this repeated motif is not clear.

(11) Transcription of the LaBelle plasmid

During the initial characterization of this plasmid Stohl et al. (1982), using nick translated ^{32}P labelled LaBelle plasmid as the probe, concluded that there was no prominent detectable plasmid-specific transcript, since they detected only very weak hybridization to RNA species of apparent sizes from 2 to 4 Kb. However, the presence of the long ORF suggests that the information on the plasmid is expressed. To facilitate detection of low level transcripts, the sensitivity of detection was improved by using labelled, strand-specific RNA to perform Northern hybridizations. To achieve this, the 3.5 Kb and 0.6 Kb EcoRI-BamHI fragments of the LaBelle plasmid were cloned into Bluescribe vectors to yield plasmids pBSA15 and pBSA1, respectively (Fig. 11). The 3.5 Kb fragment contains the entire coding sequence of the 1,151 amino acid ORF.

The transcript produced from plasmid pBSA15 using T3 RNA polymerase would be complementary to an mRNA that was capable of directing the synthesis of the entire 1,151 amino acid polypeptide encoded by the long ORF of the plasmid. However, if the opposite strand is transcribed from pBSA15, by using T7 RNA polymerase, a probe which should be noncomplementary to the putative message from the ORF would be produced. Plate I shows the Northern analysis performed utilizing both of these probes on RNA isolated from mitochondria of the LaBelle plasmid containing strain (NCN47) and a non plasmid-containing standard laboratory strain (NCN10). Hybridization to bands of estimated sizes 5.8 Kb, 4.2 Kb, 3.95 Kb, 2.35 Kb and 2.15 Kb was seen in the mtRNA isolated from the plasmid containing strain, NCN47 with both the T3 and T7 riboprobes (plate I, lanes 1 and 3). The 7.0 Kb band was observed only with the T7 riboprobe (lane 3) while the 2.95 Kb, 2.8 Kb, 2.6 Kb and 1.95 Kb bands were only observed with the T3 riboprobe (lane 1). However, no hybridization was observed between the same probes and total mtRNA from the NCN10 strain (plate I, lanes 2 and 4). These results clearly indicated that the hybridization was specific to the plasmid-containing strain and thereby precluded any possibility of nonspecific hybridization to the standard mtDNA or mtRNAs. However, it was not possible to exclude the possibility that the observed hybridization was due to the presence of small amounts of LaBelle plasmid DNA which were not removed by the RNA purification procedure. To test this possibility, the T3 and T7 pBSA15 riboprobes were used to probe blots of total mtRNA isolated from the LaBelle plasmid-containing strain, which were either untreated, treated with DNase, or treated with RNase. Each of these mtRNA samples was run in three gels. One was stained with ethidium bromide and the other two were blotted and then probed with either the pBSA15 T3 or pBSA15 T7 riboprobes.

The intense bands seen in the ethidium bromide stained gel are rRNA bands (plate II). This gel clearly shows that RNase treatment was effective in degrading the mtRNA as judged by the disappearance of the mtrRNA bands (lane 2, plate II) and that DNase does not largely affect the RNA in the sample (lane 1, plate II). However, it should be pointed out that there is a slight reduction in the amount of rRNA seen in the sample that has been treated with DNase (compare lane 1 to lane 3 in plate II). However, since DNase I treatment does not appear to have increased the amount of smearing in the bands, it seemed likely that the decreased amount was due to loss of material incurred in the phenol extraction and/ precipitation step. In order to test whether this DNase treatment affected the quality of the RNA that was present in the sample, a Northern blot was performed using a probe for cytochrome c oxidase subunit 1. The results are shown in plate III. Upon DNase treatment, the intensity of the COI band has decreased and possibly there is slight increase in smearing of RNA. The slight reduction in amount of RNA following DNase I treatment probably occurs mostly as a result of the precipitation step following DNase I treatment; however, a small amount of RNA may also be degraded due to residual RNase contamination in the DNase I enzyme. The other samples, not treated with DNase I, shown on plates I, II, III and IV were not precipitated prior to loading on the gel and these samples neither show a reduction in amount of RNA nor is the quality of RNA affected. In subsequent experiments, DNase I was digested with Proteinase K (1 mg/ml) for 1 hour in the presence of 10 mm $CaCl_2$ before using it on the RNA samples. Under these conditions the RNase activity is digested by Proteinase K but DNase activity is unaffected (Tullis and Rubin, 1980). The samples so treated were phenol extracted and ethanol precipitated before visualization in an agarose gel and similar reduction in the rRNA band intensity was observed even after using DNase I which had been so prepared. Therefore, this suggests that the reduction in the amount of RNA is due to loss of RNA in the phenol extraction and the precipitation step following DNase I treatment.

When the pBSA15 T3 riboprobe was used with samples that had been treated with DNase I, hybridization to the 6.8 Kb, 5.4 Kb, 4.2 Kb, 3.1 Kb, 2.3 Kb, 2.05 Kb, 1.3 Kb, 1.25 Kb, 1.15 Kb and 1.075 Kb bands was almost as strong as in the untreated mtRNA sample (plate IV; panel A, lanes 1 and 2). This indicates that most of the hybridization was due to complementary RNA. Since certain bands (corresponding to estimated sizes of 30.0 Kb, 22.0 Kb, 17.5 Kb, 12.5 Kb and 9.6 Kb) disappeared as a result of the DNase treatment, these were likely due to residual amounts of plasmid DNA, probably present as oligomers of increasing size in the preparation (Stohl et al., 1982). This is supported by the results obtained when hybridization was done to samples which had been treated with RNase A. Samples treated with RNase contain little or no RNA (lane 2, plate II). Virtually no hybridization was observed in these cases (lane 3, panel A; plate IV) due to digestion of complementary RNA. Faint bands appeared in the high molecular weight region (lane 3, plate IV A) which are probably due to residual plasmid DNA in the preparation. The decreased amount compared to lane 1 is probably due to losses incurred during phenol extraction and/ precipitation; and/ or residual DNase activity in the RNase preparation. The 5.4 Kb band is (lane 1, panel A; plate IV) not clear due to an artifact of the blotting procedure. The 1.15 Kb and the

1.075 Kb bands are very faint in lane 2 of this blot (panel A; plate IV). This could be accounted for by loss of some RNA during the precipitation step following DNase I treatment as explained above.

On the other hand, despite the fact that the samples which were treated with RNase contain little or no RNA (lane 2, plate II), hybridization was observed with these samples using the pBSA15 T7 riboprobe (plate IV, panel B- lane 1). However, no hybridization was seen using the same riboprobe in the lanes containing LaBelle mtRNA which had been treated with DNase (plate IV, panel B- lane 2). Thus, the 16.0 Kb, 7.0 Kb and 6.0 Kb bands seen in plate IV, panel B, are due to hybridization with residual plasmid DNA. The 16.0 Kb band is not seen in lane 3 (panel B of plate IV) of this blot due to an artifact of the blotting procedure which is evident as a sudden break in the background hybridization in the top part of lane 3. These experiments cannot explain the origin of the 4.5 Kb, 4.2 Kb, 3.1 Kb, 2.5 Kb and 2.2 Kb bands seen in panel B, plate IV. If these are DNA bands then they should be present in both lanes 1 and 3. If they represent RNA species they should be seen in lane 2 as well. Nonetheless, the total absence of hybridization in the DNase I treated sample (plate IV; panel B, lane 2) strongly suggests that the bottom strand of the LaBelle plasmid is not transcribed and that the hybridization observed was due to a small amount of contaminating plasmid DNA in the total mtRNA isolated.

Taken together, these results indicate that the upper strand (i.e. that which is complementary to the pBSA15 T3 riboprobe) of the LaBelle mitochondrial plasmid (Fig. 4) containing the long ORF, is transcribed in the mitochondria of logarithmically growing mycelia of the LaBelle-1b strain of N. intermedia.

(12) Sizes of the LaBelle transcripts

The sizes of the LaBelle plasmid specific transcripts were calculated using rabbit reticulocyte rRNA (28S rRNA = 5 Kb and 18S rRNA = 2 Kb) and yeast rRNA (28S rRNA = 3.75 Kb and 18S rRNA = 2 Kb) as size standards. Rabbit 18S rRNA has been shown to be 1.858 Kb by direct RNA sequencing and the yeast 18S rRNA sequence was deduced to be 1.789 Kb by sequencing the cloned 18S rRNA gene (Connaughton *et al*., 1984; Rubtsov *et al*., 1980). *Neurospora* mtrRNA was not used as a size standard for several reasons. Initially the size of the COI transcript was estimated to be 6.8 Kb by sizing against *S. cerevisiae* markers (Burger *et al*., 1982). Subsequently the COI transcript was determined to be precisely 5.0 Kb by SI mapping experiments (Burger and Werner, 1985). Under the experimental conditions used in this study the size of the COI transcript was estimated to be 6.5 Kb using rabbit reticulocyte and yeast rRNAs as size standards. Furthermore, the 25S rRNA (also referred to as 24S in the literature, Burger and Werner, 1985) and 19S rRNA (also referred to as 17S, Burger and Werner, 1985) from the mitochondria of *N. crassa* 74 OR23-1A strain have been used as 3.3 Kb and 2.0 Kb standards respectively (Akins and Lambowitz, 1984; Burger and Werner, 1985). However, in the gel system used in the present study, both the 25S rRNA and 19S rRNA from *N. crassa* 74A (NCN10) and *N. intermedia* strain LaBelle-1b (NCN47) consistently migrate as if they are larger than the 3.75 and 2.0 Kb size standards such that their estimated sizes would be 4.0 Kb and 2.2 Kb respectively (plate II compare lanes 1, 3, 4 to lanes 5 and 6). In the same gel system the 2.0 Kb 18S rRNAs from both rabbit and yeast consistently show similar electrophoretic mobilities. In view of all these anomalies it should be borne in mind that the transcript sizes estimated in this study may be inaccurate.

As shown in panel A of plate IV, there are three transcripts whose estimated sizes are 6.8 Kb, 5.4 Kb and 4.2 Kb, which are respectively, 2.73, 1.33 and 0.13 Kb larger than the size of the plasmid (4.07 Kb) itself. A somewhat similar situation has been described for the Varkud mitochondrial plasmid. This plasmid is transcribed to give a full-length linear RNA plus a transcript which is 5.0 Kb long; 1.2 Kb longer than the monomer plasmid (Akins *et al*., 1986). Recent observations suggest that in Varkud, the longer-than-unit-length transcript results from addition of part of the small mtrRNA to the 5' end of this transcript. Whether the addition of this small mtrRNA takes place during transcription or post-transcriptionally is not yet clear (Akins *et al*., Submitted). Transcripts that are slightly longer than unit length have also been observed in the Mauriceville mitochondrial plasmid (Nargang *et al*., 1984). It is possible an analogous situation exists in LaBelle and it would be of interest to do Northern experiments using a variety of probes including mtrRNA to check out this possibility.

Other LaBelle specific transcripts were estimated to be 3.1 Kb, 2.3 Kb, 2.05 Kb, 1.3 Kb, 1.25 Kb, 1.15 Kb and 1.075 Kb in length. The longest ORF in the LaBelle plasmid is 3,456 nucleotides long, which could be encoded by the 4.2 Kb or any longer transcript. Different batches of mtRNA preparations were used to perform experiments shown in plates I and IV. Eventhough the quality of total mtRNA as judged by the intensity of mtrRNA bands was consistently good, bands of different sizes are seen when Northerns are done on mtRNA preparations from different batches. At present all these factors make it difficult to say if these smaller transcripts are actually processed from a longer precursor transcript, are independently transcribed, or are degradation products (Mott *et al.*, 1985; Birnstiel *et al.*, 1985).

It must be reemphasized that although the results of the Northern analysis presented in this study are very preliminary they tend to indicate that the correct strand of the plasmid is transcribed to allow expression of some or all of the long ORF. Precise mapping of the 5' and 3' ends of these transcripts should be done before any definitive statements regarding their size and significance can be made.

Previous SDS gel analysis of mitochondrially synthesized proteins showed no major bands that could be clearly assigned to the LaBelle plasmid (Stohl *et al.*, 1982). However, the results obtained during this study suggest that the long ORF of the plasmid is represented in transcripts of the plasmid. Therefore, it is possible that the plasmid-encoded protein may be present in low amounts or it may comigrate with another major mitochondrial translational product. The results observed in the Northern hybridization experiments tend to favour the former view. That is, when 3.0 μg of total mtRNA isolated from the LaBelle-containing strain was present per lane and a single-stranded complementary RNA probe of 3 x 10^7 cpm was used, a minimum of 15 hours exposure time was required for all the LaBelle-specific transcripts to be detected in the autoradiogram. On the other hand when 4.5 μg of total mtRNA preparation from the LaBelle-containing strain was probed with 1 x 10^7 cpm of double stranded DNA probe for single copy mitochondrial gene cytochrome *c* oxidase subunit I, it took only 30 minutes of exposure to detect the cytochrome c oxidase subunit

I mRNA. The single stranded RNA probes are much more sensitive than the double stranded DNA probes. Therefore, a relative comparison of these exposure times seems to indicate that the LaBelle specific message in the total mtRNA preparation from the LaBelle-containing strain is not very abundant.

Due to the presence of several AUG codons in the long ORF, it is not clear how long any putative LaBelle specific protein might be. Antibodies could be raised against the synthetic oligopeptides made from the predicted antigenic regions of the putative polypeptide. These antibodies could be used to detect the LaBelle specific polypeptide from mitochondrial translation products by immuno-precipitation. A purified protein could be assayed to determine its size and to establish if this protein is indeed a RNA polymerase.

II DNA sequence analysis of homologous regions in Neurospora mtDNA and LaBelle mitochondrial plasmid DNA

It was previously reported that the LaBelle mitochondrial plasmid contained a region which hybridized weakly to the EcoRI-6 fragment of *N. crassa* mtDNA (Stohl *et al.*, 1982). Additional DNA-DNA hybridization experiments indicated that the homology between the LaBelle mitochondrial plasmid and the EcoRI-6 fragment of *Neurospora* mtDNA was predominantly confined to an internal KpnI fragment (shown in Fig. 12) of the EcoRI-6 fragment of strain 74*A* mtDNA (R. Akins, personal communication). The DNA sequence of the corresponding KpnI fragment from *N. intermedia* strain LaBelle-1b was determined and found to be 854 bp in length. Surprisingly, upon comparison of the plasmid sequence to that of the KpnI fragment, the homology between the two was found to extend completely to one end of the KpnI fragment (Fig. 13). Therefore, it was necessary to determine the DNA sequence of the region past this KpnI site to ensure that the entire region of homology was found. Thus, the sequence through the KpnI site towards the EcoRI site at the junction of the EcoRI-4 and 6 fragments was determined. Figs. 14 and 15 show the sequencing strategy and the DNA sequence, respectively, of the EcoRI-6 fragment of *N. intermedia* strain LaBelle-1b mtDNA.

Computer analysis of the DNA sequences of the mitochondrial plasmid and the homologous region from the EcoRI-6 fragment of *N. intermedia* mtDNA revealed several interesting features. The region of homology between the mitochondrial plasmid and the mtDNA may be aligned into a contiguous region containing blocks of high homology interspersed with blocks where homology is interrupted by frequent mismatches. There are also locations where "loopouts" (i.e. long regions of unmatched sequences) occur in one or the other sequence as shown in Figs. 16 and 17. To facilitate discussion of the homologous regions, each DNA sequence has been numbered. As in Section I(2) the first nucleotide in the unique EcoRI site of the LaBelle plasmid has been arbitrarily assigned position 1 of the plasmid. Position 1 of the EcoRI-6 fragment of mtDNA is arbitrarily defined as the first nucleotide in the EcoRI site at the EcoRI-4, 6 junction. Disregarding the numerous mismatches and the loopouts that occur within the region, the homology starts at nucleotide position 2,898 in the LaBelle plasmid and continues up to nucleotide position 3,982. In the EcoRI-6 fragment of mtDNA from *N. intermedia* LaBelle-1b, the region of homology starts at nucleotide position 50 and continues to nucleotide position 1,359. Thus, a region of 1,085 bp of the LaBelle mitochondrial plasmid contains homology to a region of 1,310 bp of the EcoRI-6 fragment of *N. intermedia* strain LaBelle-1b mtDNA.

The two sequences contain regions which are 10 bp or longer that have no DNA sequences corresponding to them in the other sequence. These are considered to be "loopout" regions. These loopout regions could be visualized as the result of the insertion or deletion of DNA sequences in one of the two DNA sequences. As defined above, there are a total of seven of these loopout regions (Table 12, Figs. 16 and 17). As shown in Table 12, six of these seven are the result of insertion of sequences in the EcoRI-6 fragment as compared to the LaBelle mitochondrial plasmid (alternatively they could be deletions from the LaBelle mitochondrial plasmid). The last loopout region, starting at position 3,902 of the plasmid, is the result of insertion of DNA sequence into the LaBelle mitochondrial plasmid DNA (or deletion from EcoRI-6). Two of these loopouts in the EcoRI-6 fragment contain palindromic sequences (Fig. 17 and Table 12). Several other palindromic sequences that are not within loopout regions are also seen in the EcoRI-6 fragment as shown in Table 13 and Fig. 18. It is noteworthy that most of the sequences which could make hairpin loops are present in the EcoRI-6 fragment. In fact, the LaBelle mitochondrial plasmid has only one palindromic sequence in the region of homology which is found at nucleotides 3,612 to 3,672. This region has the capacity to form a 61 bp stem loop structure and is the most stable hairpin loop that can be made in the entire LaBelle plasmid, as discussed before (section 6). Another interesting feature of the loopout regions is that three of the six sequences that are inserted in EcoRI-6, are highly G+C rich. The 33 bp insertion from nucleotide positions 244 to 276 is 91% G+C, the 33 bp insertion from nucleotide positions 493 to 525 is 64% G+C, and the 28 bp insertion from nucleotide positions 567 to 594 is 89% G+C. This is in contrast to the overall region (including these above mentioned regions) which has a G+C content of 36%. These G+C rich insertions are confined to the part of EcoRI-6 which overlaps with the carboxy terminal end of the LaBelle open reading frame (ORF) (See below). All the other insertion sequences in EcoRI-6 or the LaBelle mitochondrial plasmid are A+T rich with their G+C content ranging from 12% to 46% (Table 12).

As described earlier, the LaBelle mitochondrial plasmid contains a very long ORF beginning with the ATG codon at position 59-61 of the plasmid and extending to a stop codon at positions 3,512-3,514 (Fig. 3 and 5). Thus, 617 nucleotides at the 3' end of this long ORF are homologous to the EcoRI-6 fragment. However, the reading frame present in the plasmid ORF is not conserved in the EcoRI-6 fragment and the significance of the homology is not clear.

To determine whether divergence in the homologous sequence was greater between the plasmid and *N. intermedia* mtDNA or between the mtDNAs of different species of *Neurospora*, the sequence of the EcoRI-6 mtDNA fragment of *N. crassa* strain 74*A* was determined. This strain does not contain a mitochondrial plasmid. Fig. 19 shows the sequencing strategy used for the EcoRI-6 fragment of *N. crassa* mtDNA. The DNA sequence of this region is shown in Fig. 20.

Comparison of the DNA sequences from the two mtDNA's revealed almost total conservation (Fig. 13). This comparison was restricted to the region of the DNA sequence where both the DNA strands had been fully sequenced in the two above mentioned mtDNA fragments (A total of 942 bp). The only differences in the two sequences are: 1) a single insertion of 139 bp in the *N. intermedia* strain relative to *N. crassa* strain after nucleotide position 415 (this corresponds to nucleotide position 695 to nucleotide position 833 of *N. intermedia*, as shown in Figs. 13 and 17). 2) the *N. intermedia* strain has only 8 G's (nucleotide position 492 to 499) as compared to 9 G's in the corresponding *N. crassa* sequence (nucleotide position 212 to 220) as shown in Fig. 13.

(1) PstI palindromes

As shown in Figs. 14, 15, 19 and 20, the EcoRI-6 fragments from both *Neurospora* species studied contain 2 perfect palindromic PstI sequences. In *Neurospora* mtDNA these elements are found flanking tRNA and other genes (Yin _et al_., 1981; Burke and RajBhandary, 1982; Browning and RajBhandary, 1982; Macino and Morelli, 1983). Computer analysis of the DNA sequence located between these two perfect PstI palindromic sequences did not reveal the presence of any tRNA gene. Initially PstI palindromes were thought to be processing signals (Yin _et al_., 1981; Yin _et al_., 1982; Burke and RajBhandary, 1982; Browning and RajBhandary, 1982; Macino and Morelli, 1983) in _Neurospora_ mitochondrial transcripts. However, Breitenberger _et al_., (1985) and Burger _et al_., (1985) have demonstrated that PstI palindromes do not act as processing signals because they are retained in several processed mature RNAs . In fact, RNA processing in _N. crassa_ mitochondria has been shown to be mediated by tRNA sequences, which also act as primary processing signals (Breitenberger _et al_., 1985) as in the case of mammalian (Ojala _et al_., 1981), and possibly amphibian and insect mitochondria (Wong _et al_., 1983; de Bruijn, 1983).

III Heterokaryon analysis

Different N. intermedia strains contain different mitochondrial plasmids (Stohl et al., 1982; Nargang, 1986). At the time this work was begun different mitochondrial plasmids had not yet been reported to occur simultaneously in the mitochondria of one strain.

It was thought that if heterokaryons could be forced between appropriate mitochondrial plasmid containing auxotrophic strains there would be three possible outcomes: 1) the mitochondrial plasmids will recombine, 2) one mitochondrial plasmid will be suppressive toward the other, or 3) the plasmids will simply coexist without interaction.

To distinguish between these possibilities two different mitochondrial plasmid containing strains were chosen, N. intermedia strain LaBelle-1b containing the 4,070 bp LaBelle mitochondrial plasmid, and N. intermedia strain Varkud-1c containing the 3,675 bp Varkud mitochondrial plasmid. These plasmids have distinct restriction patterns and are not homologous. Thus, the origin of any recombinant plasmid arising from recombination of these original plasmids could be easily distingished.

Since both the plasmid containing strains were wild-type prototrophic strains, it was necessary to construct auxotrophs to insure that heterokaryons were forced. Auxotrophic strains of N. intermedia strain LaBelle-1b were constructed by crossing NCN47 (LaBelle-1b, a FGSC # 1940) as the protoperithecial (female) parent to NCN55 (N. intermedia his-A, A FGSC# 3391) as the male parent. In general, NCN47 did not behave efficiently as the protoperithecial parent. Of the few ascospores that were discharged, 2 out of 25 isolates, named L5 and L11, were both hisA; A and contained the LaBelle mitochondrial plasmid.

To obtain auxotrophic strains of N. intermedia strain Varkud -1c (FGSC # 1823, A, called NCN50) several crosses were performed. NCN50 was used as the female parent and crossed to different auxotrophic strains of N. intermedia. Several auxotrophic progeny containing the Varkud plasmid were obtained.

Separate heterokaryons were forced on minimal media between LaBelle A; his A auxotrophs and various Varkud A auxotrophs. Ideally, heterokaryons are formed within 12 to 24 hours and these grow at the wild-type rate thereafter (Davis and de Serres, 1970). However, no stable heterokaryons could be obtained between the plasmid-containing strains. Typically, if heterokaryons formed, they began to grow only after 72 to 96 hours. Subsequent growth was much slower than that of wild-type strains. Darkening of the agar media, a typical manifestation of heterokaryon incompatibility (Davis and de Serres, 1970), usually followed.

Strains carrying nonallelic mutations may often fail to complement because of differences in heterokaryon-compatibility genes (Mylyk, 1975). At least ten heterokaryon-incompatibility genes are known in Neurospora crassa (Mylyk, 1975). In the initial stages of this work, no attempt was made to isolate compatible strains. Just as crosses designed to produce compatible strains were about to be started, it was discovered that wild strains carrying different plasmids did exist (Taylor et al., 1985), and the study was not pursued.

CHAPTER IV

CONCLUSIONS

(1) The LaBelle plasmid

Prolonged exposures of autoradiograms using ^{32}P labelled LaBelle plasmid as probe showed heterogeneous RNAs from *N. intermedia* strain LaBelle-1b, having lengths 2 to 4 Kb (Stohl *et al*., 1982). The significance of the weak hybridization was unclear to the authors though they speculated that it could be due to low concentration of the transcript in the mtRNA preparation, unusually high AT-contents, a small portion of undigested plasmid DNA, or weak association between the plasmid DNA probes and normal mtRNA species. The results of DNA sequence analysis show that the plasmid is relatively AT rich (65.1%). However, Northern analysis using strand specific RNA-probes clearly shows (plate I) that there is no association between plasmid-specific probes and normal mtRNA species (plate I). Results shown in plate IV suggest that only the upper strand (Fig. 4) of the LaBelle mitochondrial plasmid, containing the long ORF, is transcribed.

Several different lines of evidence suggest that the LaBelle plasmid is related to mtDNA. The codon usage of the plasmid encoded ORF shows a bias approaching that of both fungal mtDNA genes and mitochondrial intronic ORFs in that only 29% of codons end in C or G. *Neurospora* mtDNA codons end in C or G only about 17% of the time, while *Neurospora* nuclear gene codons end in C or G about 70% of the time. The LaBelle plasmid ORF uses several codons such as UUA (leu), AUA (ile), and UGU (cys) that are used infrequently or not at all in *Neurospora* nuclear genes (Schectman and Yanofsky, 1983; Woudt *et al*., 1983; Rutledge, 1984; Huiet and Giles, 1986; Hager *et al.*, 1986; Kinnaird and Fincham, 1983; Arends and Sebald, 1984; Viebrock *et al.*, 1982; Legerton and Yanofsky, 1985; Newbury *et al.*, 1986). In addition, the LaBelle plasmid ORF uses certain distinctive codons (CGN for arginine, AAG for lysine, UGG for tryptophan, ACC and ACG for threonine, and CUC for leucine) that are used in mtDNA introns but are used infrequently or not at all in standard mtDNA genes. Thus, it seems that the LaBelle mitochondrial plasmid bears a relationship to fungal mtDNA intronic ORFs. However, it should be noted that the plasmid has neither the consensus sequence elements common to group I nor group II mtDNA intronic elements (Michel *et al.*, 1982; Michel and Dujon, 1983) nor does it seem to encode any intronic maturase (Waring *et al.*, 1982; Hensgens *et al.*, 1983).

Suggestions have been made that mitochondrial introns have arisen from insertion of an invading mobile genetic element (Dujon *et al.*, 1987). It was also suggested that the two plasmids Varkud and Mauriceville might belong to a class of mobile elements that are related to the progenitors of mtDNA introns (Nargang *et al.*, 1984; Lambowitz *et al.*, 1985). Although the actual transposition of introns has not yet been demonstrated, the rl intron of the 21S rRNA gene of *S. cerevisiae* ω^+ strains have been shown to encode a specific transposase. This ω transposase shows a double-strand endonuclease activity which generates a 4-bp staggered cut with 3'OH overhangs within a specific sequence of the 21S rRNA gene of ω^- strains and causes duplicative transposition of the rl intron into this sequence (Dujon *et al.*, 1987). This results in the conversion of an ω^- strain to ω^+. An intron that is closely related to the rl intron of

S. cerevisiae and *Kluyveromyces thermotolerans* has been shown to be present within the ATPase subunit 9 gene of *Kluyveromyces fragilis*. So far insertion of this rl related intron within the ATPase subunit 9 gene of *K. fragilis* has not been directly demonstrated but it has been suggested that this is the end result of a transposition event that took place in the past (Dujon *et al*., 1987). Evidence supporting the possibility that Mauriceville and Varkud might have similar capabilities has been obtained (Akins *et al*., 1986) by the demonstration that suppressive variant plasmids are generated by a mechanism that includes a reverse transcription step, that the plasmids can integrate into mtDNA in a manner consistent with integration via an RNA intermediate, and that some of the integration events may have been promoted by a site specific DNA endonuclease activity.

The LaBelle plasmid also seems to be related to mitochondrial introns, however, the arguments for the Mauriceville and Varkud plasmids being progenitors of present day introns cannot be directly applied to LaBelle because of several differences between them. Based on DNA-DNA hybridization LaBelle was placed in a homology group (group 3) separate from that of Mauriceville and Varkud (group 2) and other known *Neurospora* mitochondrial plasmids (Natvig *et al*., 1984). The analysis of the LaBelle plasmid DNA sequence further corroborates the differences between *Neurospora* mitochondrial group 3 and group 2 plasmids. The LaBelle plasmid does not contain the conserved sequence elements characteristic of either group I or group II fungal mitochondrial introns. Furthermore Mauriceville and Varkud give rise to abundant discreet transcripts while LaBelle gives rise to a number of transcripts ranging in size from 6.8 Kb to 1.0 Kb. The LaBelle plasmid contains several palindromic sequences but does not contain PstI-palindromes, which are found to be flanking the genes in *Neurospora* mtDNA and in the Mauriceville and Varkud plasmids (Yin *et al*., 1981; Nargang *et al*., 1983, 1984). The putative LaBelle plasmid-encoded proteins showed no homology to the putative polypeptide encoded by the Mauriceville ORF or any other reverse transcriptases. However, the 1,151 amino acid long polypeptide encoded by the long ORF of the LaBelle plasmid shows homology to various RNA polymerases.

A distinctive feature of organelle introns in lower eukaryotes is their coding capacity, for they often contain long internal ORFs (Dujon *et al*., 1987). A number of functions have been postulated for the putative translation products of these ORFs. Many of these proteins were postulated to be "maturases" and later several of them were shown by genetic analysis to participate in the excision-splicing process (Lazowska *et al*., 1980; Carignani *et al*., 1983). In *S. cerevisiae* a group I intron has been shown to encode a maturase (Lazowska *et al*., 1980) and the maturase encoded by the first intron of the gene for the subunit I of the cytochrome oxidase in *S. cerevisiae* provides an example of such an enzyme from the group II introns (Carignani *et al*., 1983). The protein encoded by the ω^+ intron was shown to be active in gene conversion of ω^- strains to ω^+, thus spreading the ω^+ intron into a mitochondrial gene (Jacquier and Dujon, 1985). Macreadie *et al*., (1985) demonstrated that transposition of this intron actually requires the protein encoded by that intron itself and that this protein is a site specific endonuclease. Michel and Lang, (1985) have shown that the proteins encoded by four mitochondrial group II introns are related to several RNA-dependant polymerases that have viral and transposable element origins.

The results presented in this study show that the 4.07 Kb LaBelle mitochondrial plasmid contains a 3,456 nucleotide long ORF. Codon usage in this ORF suggests that it may also be related to mtDNA introns. One of the several plasmid-specific transcripts that are transcribed _in vivo_ has the potential to be translated into the putative protein product. The existence of such a long ORF argues that it probably encodes a function that is under selective pressure. A similar argument for conservation of function was applied to the ORF sequence found and conserved in the Varkud and Mauriceville mitochondrial plasmids (Nargang , 1986). Although no plasmid-specific proteins have been demonstrated in any of these cases, much circumstantial evidence exists to suggest that the proteins of the Mauriceville and the Varkud plasmids might have reverse transcriptase and/ or DNA endonuclease activities associated with them (Akins _et_ _al_., 1986; Akins _et_ _al_., Submitted).

What is the function of this putative ORF encoded protein in LaBelle ? One possible function is that of a plasmid-encoded RNA polymerase. This tentative function is assigned based on the similarities found with different RNA polymerase proteins and parts of the 1,151 amino acid polypeptide specified by the long ORF of the LaBelle plasmid. It should be noted that neither is any single region of similarity very long, nor are the levels of identities very high. However, these are the highest levels of similarity that have been found between the LaBelle long ORF and any sequences present in the Gene Bank database. It is therefore proposed that all of these regions of similarity considered together do suggest a relationship between the

ORF and RNA polymerase.

(2) Region of homology to mtDNA

A region of 1,085 bp of the LaBelle mitochondrial plasmid has been shown to be homologous to a region of 1,310 bp of the EcoRI-6 fragment of *N. intermedia* strain LaBelle-1b mtDNA. None of the other *Neurospora* nuclear or mitochondrial plasmids that have been studied to date have any homology to the mtDNA (Collins *et al*., 1981; Stohl *et al*., 1982; Natvig *et al*., 1984; Taylor *et al*., 1985; Bertrand *et al*., 1986; Bertrand, 1987). This region of homology contains blocks of high homology interspersed with blocks where homology is interrupted by frequent mismatches or loopout regions. Most of the loopout and palindromic sequences are present in the mtDNA when compared with the LaBelle plasmid. The latter has only one loopout region and only one palindromic sequence. The loopouts in the EcoRI-6 fragment of mtDNA are extremely G+C rich.

In yeast, mtDNA G+C clusters are found embedded in A+T stretches (Cosson and Tzagoloff, 1979; Gaillard and Bernardi, 1979). It has been demonstrated that the directly repeated sequences within A+T stretches and G+C clusters are used as sites for the excision of the defective genomes of spontaneous petite mutants in yeast (Baldacci *et al*., 1980; Gaillard *et al*., 1980) and also play a role in site-specific recombination events (Fonty *et al*., 1978). In this context, it is conceivable that G+C rich segments represent sites where such events can take place in the region of homology.

The region that can form the most thermodynamically stable secondary structure in the LaBelle mitochondrial plasmid begins at position 3,612. The region is partially homologous to a region in EcoRI-6. The secondary structures possible in EcoRI-6 are similar, but not identical, to those found in the mitochondrial plasmid (Fig. 22). It has been suggested that the *N. crassa* mtDNA EcoRI-4/6 fragment junction might contain a mtDNA replication origin (Collins and Lambowitz, 1981). The mtDNA replication origin regions of yeast, *Aspergillus amstelodami*, rat, mouse, bovine, and Hela cells have all been shown to have the potential for forming characteristic hairpin loop structures (Zamaroczy *et al*., 1984). One of the *Nicotiana tabacum* chloroplast autonomously replicating sequences has also been shown to have the potential to form a stemloop structure (Ohtani *et al*.,1984). Therefore, it seems likely that the origin of replication of *Neurospora* mtDNA should also be capable of forming a hairpin loop structure and it is conceivable that the secondary structures described above might fulfill that function in both the mitochondrial plasmid and the mtDNA. The high conservation of DNA sequences in the EcoRI-6 fragments of the two strains also suggests that selective pressure may be acting on this region despite the fact that the region contains no known genetic information. The presence of the mtDNA origin of replication in this EcoRI-6 fragment would account for this selective pressure.

The homology between the LaBelle mitochondrial plasmid and the EcoRI-6 fragment of mtDNA could be explained in three ways: 1) the mitochondrial plasmid has acquired the sequence from the mtDNA of the host organism; 2) the mtDNA of the host has acquired the sequence from the mitochondrial plasmid; or 3) the mitochondrial plasmid and the mtDNA have independently evolved similar sequences in response to a common selective pressure.

Comparison of the mtDNA sequences from the two *Neurospora* species investigated revealed almost total conservation of sequences in the EcoRI-6 fragments. This very high conservation of sequences in the *N. intermedia* strain LaBelle-1b and the laboratory *N. crassa* strain is in agreement with the observations of Collins and Lambowitz, (1983); and Taylor *et al*., (1986), who have demonstrated that the mtDNAs of different *Neurospora* strains are characterized chiefly by polymorphic length differences. On the other hand, base pair changes, scored on the basis of the occurrence of restriction site alterations between strains, occur rarely.

In view of the fact that the LaBelle plasmid sequence contains much greater sequence divergence than do the two mtDNAs, there are three main possibilities for the homology between plasmid and mtDNA. First, it is possible that the plasmid existed before the separation of the two species and the divergence from the parent sequence has been occurring for some time. Second, it is possible that *N. intermedia* acquired the plasmid after the species separated. If this were the case, the region of homology on the LaBelle plasmid might represent an example of convergent evolution. This would be an especially attractive hypothesis if the possibility that the region of homology may contain a replication origin could be shown to be true. That is, a foreign genetic element might have increased its fitness within mitochondria by evolving a replication origin similar to that of the host mtDNA's. Third, it is possible that irrespective of whether the plasmid existed before the species diverged or *N. intermedia* acquired the plasmid after the species separated, the greater sequence divergence of the LaBelle plasmid compared to the two mtDNAs is simply a reflection of the possibility that the plasmid can afford to suffer mutations in the region of homology but the mtDNA can not.

The region of homology in the LaBelle mitochondrial plasmid may contain a replication of origin. Since this study clearly demonstrates that only the EcoRI-6 fragment of mtDNA is part of the homologous fragment, it seems to delimit mitochondrial EcoRI-6 fragment as the one which contains most, if not all, of the mitochondrial origin of replication in Neurospora. This conclusion would also explain the observation that, though most mitochondrial deletion mutants of Neurospora seem to retain EcoRI-4/6 junction, some mutants do exist which retain only the EcoRI-6 fragment (Bertrand et al., 1980). This conclusion though is contradictory to the suggestion made by de Vries et al., (1981). They had concluded that their preliminary experiments indicated that a sequence homologous to the yeast mitochondrial origin of replication is present in the EcoRI-4 mtDNA fragment.

It has been speculated that the Mauriceville plasmid directs the synthesis of a protein product required for the propagation or expression of the plasmid (Nargang et al., 1984; Akins et al., 1986). It is conceivable that the LaBelle plasmid may encode a similar function. If the assumptions that the 1,151 amino acid polypeptide encoded by the long ORF of the LaBelle plasmid is indeed a plasmid encoded RNA polymerase and that the propagation (that is, replication) of the plasmid is somehow dependent on the transcription of the plasmid, are correct, then it is possible that the 1,151 amino acid long protein encoded by the long ORF of the LaBelle plasmid is involved in the replication of this plasmid. Such a protein probably would not be required in copious amounts thus explaining the difficulty in detecting this protein, in labelled mitochondrial translational products.

TABLES

Table 1. Plasmids found in *Neurospora* species

Species	Strain and/or isolation point [FGSC No. or Perkins (P) Collection No.]	Structure	Size (kb)	Reference	Homology group
N. tetrasperma	Lihue-1v2, Hawaii (2509)	circle	5.0	Natvig *et al.*, 1984	1
	Hanalei-1bv1, Hawaii (2510)	circle	5.0	" " "	1
	Waimea Falls-2, Hawaii (3296)	circle	5.0	" " "	1
	Surinam, 85A (1270)	circle	5.0	" " "	1
N. crassa	Roanoke, Louisiana (P516)	circle	5.2	Taylor *et al.*, 1985 Taylor *et al.*, 1986	1
	" " (P518)			" " "	1
	" " (P521)			" " "	1
	" " (P524)			" " "	1
	" " (P526)			" " "	1
	Houma, Louisiana (P497)			" " "	1
	Welsh, Louisiana (P515)			" " "	1
N. intermedia	Fiji N6-6 (435)	circle	5.2	Stohl *et al.*, 1982	1
	LaBelle-1b, Florida (1940 or P405)	circle	4.1	" " "	3
	Varkud-1c, India (1823)	circle	3.8	Nargang, F.E., 1986	2
N. crassa	Mauriceville-1c, Texas (2225)	circle	3.6	Collins *et al.*, 1981	2
N. intermedia	Kalilo Kal (P561)	linear	9.0	Griffiths & Bertrand, 1984; Bertrand *et al.*, 1985; Bertrand *et al.*, 1986;†;#	na

Table 1. continued:

Species	Strain and/or isolation point (FGSC No. or Perkins Collection No.)	Structure	Size (kb)	Reference	Homology group
N. intermedia	Kauai, Hawaii Kal (P608)			Rieck *et al.*, 1982;*	na
	" " Kal (P804)			" " "	"
	" " Kal (2363)			" " "	"
	" " Kal (2360)			Rieck *et al.*, 1982; Bertrand *et al.*, 1986;#	"
	" " Kal (2360 his)			" " "	"
	" " Kal (P573)			Griffiths & Bertrand, 1984	"
	" " Kal (P604)			" " "	"
	" " Kal (2316)			" " "	"
	" " Kal (P558)			Griffiths & Bertrand, 1984, Griffiths *et al.*, 1986	"
	" " Kal (P594)			" " "	"
	" " Kal (P597)			" " "	"
	" " Kal (P599)			" " "	"
	" " Kal (P606)			" " "	"
	" " Kal (P612)			" " "	"
	" " Kal (P639)			" " "	"
	" " Kal (P791)			" " "	"
	" " Kal (P792)			" " "	"
	" " Kal (P796)			" " "	"
	" " Kal (P801)			" " "	"
	" " Kal (2366)			" " "	"

Table 1. continued:

Species	Strain and/or isolation point (FGSC No. or Perkins Collection No.)	Structure	Size (kb)	Reference	Homology group
N. intermedia	Kauai, Hawaii Kal (P564)			Griffiths & Bertrand, 1984; Bertrand et al., 1985; Bertrand et al., 1986; Griffiths et al., 1986;†;#	na
	" " Kal (P572)			" " "	"
	" " Kal (P595)			" " "	"
	" " Kal (P790)			" " "	"
	" " Kal (P802)			" " "	"
	" " Kal (2365)			" " "	"
	" " Kal (P607)			Griffiths & Bertrand, 1984; Bertrand *et al*., 1985; Griffiths *et al*., 1986;†	"
	" " Kal I18			Griffiths *et al*., 1986,	"
	Oahu, Hawaii Kal (P765)			Griffiths & Bertrand, 1984	"
	Kal (P776)			" " "	"
	Maui, Hawaii Kal (P785)			" " "	"
N. crassa	Maranhar, Aarey, India	linear	6.7	Bertrand, H. Personal Communication	"

Table 1. continued:

An updated synopsis of plasmid DNA species in fungi is available in Esser *et al.* (1986)
The Kal prefix on Hawaiian strains indicates cytoplasmically based senescence but does not imply a common mechanism.
* cytoplasmically based senescence implied (Griffith and Bertrand, 1984)
† shown to have Kal DNA insertion in mtDNA during senescence (Bertrand *et al.*, 1985)
# strains shown to have Kal DNA
na not applicable

Table 2. Base composition of *Neurospora* genomic DNA, mtDNA and mitochondrial plasmids.

	A+T content in %	G+C content in %	Reference
Genomic DNA	46.0	54.0	Villa & Storck, 1968.
mtDNA	60.0	40.0	Bernard *et al*., 1975; Terpstra *et al*., 1977.
Mauriceville	58.6	41.4	Nargang *et al*., 1984.
Varkud	58.3	41.7	Akins *et al*., Submitted
LaBelle	65.1	34.9	This study

Table 3. GC rich regions in the LaBelle mitochondrial plasmid.

	%GC richness	Nucleotide positions		Symbol in Fig. 3
		start	stop	
1.	75	3700	3722	m
2.	70	21	40	a
3.	70	1790	1809	e
4.	70	2871	2892	i
5.	70	3944	3972	o
6.	65	42	70	b
7.	65	610	630	c
8.	65	1361	1380	d
9.	65	3983	4005	p
10.	60	1829	1848	f
11.	60	2206	2224	g
12.	60	2689	2708	h
13.	60	3143	3165	j
14.	60	3293	3313	k
15.	60	3368	3387	l
16.	60	3856	3875	n

Table 4. Positions, lengths, polarity indices, molecular weights, G+C and A+T percentages, and percentage of codons ending in G/C and A/T of ORFs ending at stop codon 3512-3514 and starting at all possible preceding AUG codons as initiating codons.

Position of start codon	Length of ORF in		Polarity index in %	Ratio of basic AA*/ acidic AA*	Molecular wt. in daltons	G+C%	A+T%	Codons ending in	
	nucleotides	amino acids						G/C in %	A/T in %
59-61	3453	1151	49.1	1.34	131,727	34.5	65.5	28.67	71.32
110-112	3402	1134	48.8	1.29	129,698	34.4	65.6	28.48	71.52
281-283	3231	1077	48.8	1.27	123,262	34.8	65.2	28.88	71.12
329-331	3183	1061	48.8	1.28	121,438	34.9	65.1	28.94	71.07
389-391	3123	1041	48.9	1.28	119,085	35.1	64.9	29.01	70.99
455-457	3057	1019	48.7	1.27	116,454	35.4	64.6	29.54	70.46
530-532	2982	994	49.0	1.30	113,326	35.4	64.6	29.18	70.83
584-586	2928	976	49.1	1.30	111,316	35.4	64.6	28.76	71.21
809-811	2703	901	49.5	1.31	102,901	35.3	64.7	28.75	71.25
851-853	2661	887	49.5	1.31	101,342	35.2	64.8	28.64	71.36
962-964	2550	850	49.7	1.33	97,063	35.5	64.5	28.59	71.41
1061-1063	2451	817	49.8	1.34	93,190	35.5	64.5	28.52	71.48
1118-1120	2394	798	49.9	1.39	91,088	35.6	64.4	28.45	71.55
1163-1165	2349	783	49.7	1.42	89,428	35.5	64.5	28.35	71.65
1274-1276	2238	746	49.5	1.38	85,202	35.7	64.3	28.28	71.72
1283-1285	2229	743	49.7	1.38	84,795	35.8	64.2	28.26	71.74
1286-1288	2226	742	49.7	1.38	84,663	35.8	64.2	28.17	71.83
1313-1315	2199	733	49.8	1.38	83,663	35.7	64.3	28.10	71.90
1352-1354	2160	720	49.6	1.45	82,091	35.8	64.2	27.92	72.08
1535-1537	1977	659	49.8	1.38	75,273	35.8	64.2	28.68	71.32
1583-1585	1929	643	49.8	1.39	73,324	35.8	64.2	27.84	72.16
1586-1588	1926	642	49.8	1.39	73,192	35.8	64.2	27.72	72.27
1622-1624	1890	630	49.8	1.38	71,847	36.0	64.0	28.10	71.91
1820-1822	1692	564	50.5	1.37	64,483	36.0	64.0	28.19	71.81
1865-1867	1647	549	50.6	1.36	62,777	35.9	64.1	28.05	71.95

Table 4. continued:

Position of start codon	Length of ORF in		Polarity index in %	Ratio of basic AA*/ acidic AA*	Molecular wt. in daltons	G+C%	A+T%	Codons ending in	
	nucleotides	amino acids						G/C in %	A/T in %
1922-1924	1590	530	50.8	1.45	60,371	36.0	64.0	28.49	71.51
1946-1948	1566	522	51.0	1.45	59,532	35.9	64.1	28.16	71.84
1970-1972	1542	514	51.2	1.46	58,681	35.8	64.2	28.21	71.79
2012-2014	1500	500	51.0	1.42	57,032	35.9	64.1	28.40	71.60
2030-2032	1482	494	51.0	1.47	56,329	36.0	64.0	28.34	71.66
2183-2185	1329	443	52.6	1.54	50,506	36.1	63.9	28.67	71.33
2195-2197	1317	439	52.9	1.56	50,085	36.1	63.9	28.70	71.30
2216-2218	1296	432	53.0	1.54	49,382	35.9	64.1	28.24	71.76
2291-2293	1221	407	52.8	1.51	46,218	36.1	63.9	28.75	71.25
2348-2350	1164	388	52.1	1.43	43,843	36.2	63.8	29.12	70.88
2510-2512	1002	334	54.2	1.35	38,083	36.3	63.7	28.74	71.26
2645-2647	867	289	54.7	1.49	33,212	37.0	63.0	28.72	71.28
2816-2818	696	232	55.2	1.63	26,777	35.5	64.5	25.43	74.57

AA* Amino acids.

Table 5. Positions and lengths of open reading frames (ORFs) in the LaBelle mitochondrial plasmid

Using the *Neurospora* Mitochondrial Genetic Code

	Nucleotide positions of Start Codon	Nucleotide positions of Stop Codon	Length of ORF in amino acids
Frame 2	59-61	3512-3514	1151
Frame 3	945-947	1080-1082	45
	1827-1829	2004-2006	59
	2913-2915	3057-3059	48
	3231-3233	3348-3350	39
	3849-3851	4014-4016	55
Frame 5	2323-2321	2134-2132	63
Frame 6	3210-3208	3090-3088	40
	2706-2704	2496-2494	70
	1314-1312	960-958	118
	600-598	423-421	59
	111-109	4010-4008	57

Using the Standard Genetic Code

Frame 2	59-61	1706-1708	549
	1820-1822	1907-1909	29
	1922-1924	2168-2170	82
	2183-2185	2765-2767	194
	2816-2818	3035-3037	73

Table 6. Codon usage in the long (1,151 amino acid) open reading frame of the LaBelle mitochondrial plasmid

Amino acid	Codon	LaBelle 1,151 amino acid ORF			*N. crassa* mt genes exons[1]			*N. crassa* intronic ORFs[2]			*N. crassa* nuclear genes[3]		
		NO	%^{a}	%^{b}	NO	%^{a}	%^{b}	NO	%^{a}	%^{b}	NO	%^{a}	%^{b}
phe	UUU	31	2.69	68.9	184	5.96	70.8	119	4.75	83.2	41	0.77	21.0
	UUC	14	1.22	31.1	76	2.46	29.2	24	0.96	16.8	154	2.88	79.0
leu	UUA	38	3.30	34.2	300	9.71	66.2	178	7.11	61.2	4	0.07	0.8
	UUG	34	2.95	30.6	43	1.39	9.5	22	0.88	7.6	58	1.08	11.3
	CUU	15	1.30	13.5	50	1.62	11.0	41	1.64	14.1	124	2.32	24.1
	CUC	4	0.34	3.6	4	0.13	0.9	7	0.28	2.4	227	4.24	44.2
	CUA	11	0.95	9.9	52	1.68	11.5	36	1.44	12.4	19	0.35	3.7
	CUG	9	0.78	8.1	4	0.13	0.9	7	0.28	2.4	82	1.53	16.0
ile	AUU	38	3.30	45.2	134	4.34	38.7	98	3.92	42.2	86	1.61	28.9
	AUC	7	0.61	8.3	40	1.30	11.6	10	0.40	4.3	204	3.81	68.5
	AUA	39	3.39	46.4	172	5.57	49.7	124	4.95	53.5	8	0.15	2.7
met	AUG	38	3.30	100.0	79	2.56	100.0	34	1.36	100.0	98	1.83	100.0
val	GUU	29	2.52	36.7	87	2.82	42.2	56	2.24	43.8	97	1.81	23.1
	GUC	9	0.78	11.4	10	0.32	4.9	8	0.32	6.3	259	4.84	61.7
	GUA	24	2.08	30.4	84	2.72	40.8	54	2.16	42.1	19	0.35	4.5
	GUG	17	1.48	21.5	25	0.81	12.1	10	0.40	7.8	45	0.84	10.7
ser	UCU	20	1.74	23.0	95	3.08	32.7	73	2.92	33.6	74	1.38	20.3
	UCC	4	0.35	4.6	13	0.42	4.5	18	0.72	8.3	138	2.58	37.8
	UCA	13	1.13	14.9	56	1.81	19.2	49	1.96	22.6	14	0.26	3.8
	UCG	6	0.52	6.9	5	0.16	1.7	13	0.52	6.0	50	0.93	13.7
	AGU	34	2.95	39.1	103	3.33	35.4	53	2.12	24.4	19	0.35	5.2
	AGC	10	0.87	11.5	19	0.62	6.5	11	0.44	5.1	70	1.31	19.2

Table 6. Continued.

Amino acid	Codon	LaBelle 1,151 amino acid ORF			*N. crassa* mt genes exons[1]			*N. crassa* intronic ORFs[2]			*N. crassa* nuclear genes[3]		
		NO	%[a]	%[b]	NO	%[a]	%[b]	NO	%[a]	%[b]	NO	%[a]	%[b]
pro	CCU	16	1.39	44.4	73	2.36	67.6	49	1.96	60.5	60	1.12	23.0
	CCC	5	0.43	13.9	4	0.13	3.7	8	0.32	9.9	126	2.35	48.3
	CCA	10	0.87	27.8	26	0.84	24.1	22	0.88	27.2	36	0.67	13.8
	CCG	5	0.43	13.9	5	0.16	4.6	2	0.08	2.5	39	0.73	14.9
thr	ACU	28	2.43	50.0	70	2.27	44.3	59	2.36	55.1	70	1.31	24.1
	ACC	10	0.87	17.9	8	0.26	5.1	9	0.36	8.4	154	2.88	53.1
	ACA	14	1.22	25.0	77	2.49	48.7	31	1.24	29.0	40	0.75	13.8
	ACG	4	0.35	7.1	3	0.10	1.9	8	0.32	7.5	26	0.49	9.0
ala	GCU	26	2.26	59.1	84	2.72	46.2	49	1.96	57.6	165	3.08	30.1
	GOC	6	0.52	13.6	18	0.58	9.9	13	0.52	15.3	306	5.72	55.8
	GCA	8	0.69	18.2	64	2.07	35.2	14	0.56	16.5	32	0.60	5.8
	GCG	4	0.35	9.1	16	0.52	8.8	9	0.36	10.6	45	0.84	8.2
tyr	UAU	42	3.65	75.0	111	3.59	76.0	102	4.08	85.7	25	0.47	18.4
	UAC	14	1.22	25.0	35	1.13	24.0	17	0.68	14.3	111	2.07	81.6
his	CAU	12	1.04	80.0	52	1.68	78.8	32	1.28	88.9	31	0.58	25.4
	CAC	3	0.26	20.0	14	0.45	21.2	4	0.16	11.1	91	1.70	75.0
gln	CAA	13	1.13	65.0	58	1.88	89.2	59	2.36	89.4	34	0.64	17.8
	CAG	7	0.61	35.0	7	0.23	10.8	7	0.28	10.6	157	2.93	82.2
asn	AAU	40	3.47	78.4	116	3.76	88.6	181	7.23	84.2	25	0.47	14.8
	AAC	11	0.95	21.6	15	0.49	11.4	34	1.36	15.8	144	2.69	85.2

Table 6. Continued.

Amino acid	Codon	LaBelle 1,151 amino acid ORF			*N. crassa* *mt genes* exons[1]			*N. crassa* *intronic* ORFs[2]			*N. crassa* *nuclear* genes[3]		
		NO	$\%^a$	$\%^b$	NO	$\%^a$	$\%^b$	NO	$\%^a$	$\%^b$	NO	$\%^a$	$\%^b$
lys	AAA	92	7.99	73.0	66	2.14	88.0	199	7.95	77.1	18	0.34	6.1
	AAG	34	2.95	27.0	9	0.29	12.0	59	2.36	22.9	276	5.16	93.9
asp	GAU	72	6.25	80.0	68	2.20	85.0	93	3.72	89.4	103	1.92	39.5
	GAC	18	1.56	20.0	12	0.39	15.0	11	0.44	10.6	158	2.95	60.5
glu	GAA	36	3.13	60.0	68	2.20	76.4	75	3.00	76.5	45	0.84	13.8
	GAG	24	2.08	40.0	21	0.68	23.6	23	0.92	23.5	282	5.27	86.2
cys	UGU	15	1.30	88.2	19	0.62	86.4	19	0.76	95.0	11	0.21	17.5
	UGC	2	0.17	11.8	3	0.10	13.6	1	0.04	5.0	52	0.97	82.5
trp	UGA	6	0.52	75.0	49	1.59	94.2	26	1.04	81.3	---	----	----
	UGG	2	0.17	25.0	31	0.10	5.8	6	0.24	18.7	58	1.08	100.0
arg	CGU	24	2.08	40.0	4	0.13	5.7	15	0.60	13.5	102	1.91	31.9
	CGC	1	0.09	1.7	0	0.00	0.0	7	0.28	6.3	116	2.17	36.3
	CGA	5	0.43	8.3	2	0.06	2.9	9	0.36	8.1	26	0.49	8.1
	CGG	3	0.26	5.0	0	0.00	0.0	3	0.12	2.7	22	0.41	6.9
	AGA	19	1.64	31.7	59	1.91	84.3	60	2.40	54.1	18	0.34	5.6
	AGG	8	0.69	13.3	5	0.16	7.1	17	0.68	15.3	36	0.67	11.2
gly	GGU	42	3.65	61.8	105	3.40	50.0	63	2.52	50.0	198	3.70	46.7
	GGC	2	0.17	2.9	5	0.16	2.4	5	0.20	3.9	169	3.16	39.9
	GGA	9	0.78	13.2	74	2.40	35.2	38	1.52	30.2	36	0.67	8.5
	GGG	15	1.30	22.1	26	0.84	12.4	20	0.80	15.9	21	0.39	4.9

Table 6. Continued.

Data compiled from:

1. References: de Jonge & de Vries, 1983; Macino & Morelli, 1983; Browning & RajBhandary, 1982; Burger & Werner, 1985; Morelli & Macino, 1984; de Vries *et al*., 1986; Nelson & Macino, 1987.

2. References: Burger & Werner, 1985; Morelli & Macino, 1984; Burke & RajBhandary, 1982; Burke *et al*., 1984; Nelson & Macino, 1987.

3. References: Schectman & Yanofsky, 1983; Woudt *et al*., 1983; Kinnaird & Fincham, 1983; Arends & Sebald, 1984; Viebrock *et al*., 1982; Legerton & Yanofsky, 1985; Newbury *et al*., 1986; Rutledge, 1984; Hulet & Giles, 1986; Hager *et al*., 1986.

NO: Number of various codons found in the particular ORFs or genes.

%^{a}: Percentage of various codons found in the particular ORFs or genes.

%^{b}: Percentage of a given amino acid represented by that codon in the particular ORFs or genes.

Table 7. Percentage of codons ending in G or C and A or U

	LaBelle long ORF	Mauriceville ORF[1]	*Neurospora crassa* mt DNA exons[2]	*Neurospora crassa* mt DNA introns[3]	*Neurospora crassa* nuclear genes[4]
Codons ending in G or C	28.67%	39.14%	17.10%	17.10%	70.49%
Codons ending in A or U	71.32%	61.42%	82.94%	82.94%	29.51%

Data compiled from:

1. References: Nargang *et al*., 1984.
2. References: de Jonge & de Vries, 1983; Macino & Morelli, 1983; Browning & Raj Bhandary, 1982; Burger & Werner, 1985; Morelli & Macino, 1984; de Vries *et al*.,1986; Nelson & Macino, 1987.
3. References: Burger & Werner, 1985; Morelli & Macino, 1984; Burke & Raj Bhandary, 1982; Burke *et al*., 1984; Nelson & Macino, 1987.
4. References: Schectman & Yanofsky, 1983; Woudt *et al*., 1983: Kinnaird & Fincham, 1983; Arends & Sebald,1984; Viebrock *et al*., 1982; Legerton & Yanofsky, 1985; Newbury *et al*., 1986; Rutledge, 1984; Huiet & Giles, 1986; Hager *et al*., 1986.

Table 8. Usage of distinctive codons

Codons	Amino acid	LaBelle long ORF	*Neurospora crassa* mt DNA exons	*Neurospora crassa* mt DNA introns
		Percentage use of distinctive codons		
CGN	Arg	55.0%	8.6%	30.6%
UGG	Trp	25.0%	5.8%	18.7%
ACC+ACG	Thr	25.0%	7.0%	15.9%
CUC	Leu	3.6%	0.9%	2.4%
AAG	Lys	27.0%	12.0%	22.9%
		Ratio of codon usage		
UUU/UUC	Phe	2:1	2:1	5:1
UAU/UAC	Tyr	3:1	3:1	6:1
CAU/CAC	His	4:1	4:1	8:1

Data compiled from Table 6

Table 9. List of palindromic sequences in the LaBelle mitochondrial plasmid.

	% of correct matches in stem	Length of stem in bp	Length of central hairpin loop	Palindromic sequence position of nucleotides		ΔG in Kcal/mol	Region shown in Fig.7
				start	stop		
1.	93	15	1487	129	1645	-8.2	
2.	92	13	2637	252	2914	-4.4	
3.	74	14	7	291	330	-12.2	*
4.	94	16	3611	337	3979	-27.6	
5.	92	13	3344	476	3845	-8.0	
6.	79	15	6	707	746	-20.8	*
7.	93	14	240	954	1222	-9.4	
8.	83	20	1336	1305	2683	-20.4	
9.	92	12	2005	1532	3561	-9.8	
10.	92	13	1323	1561	2909	-7.2	
11.	61	20	0	1712	1764	-11.6	*
12.	92	12	927	2584	3535	-13.4	
13.	89	17	932	2606	3572	-8.8	
14.	100	8	6	2700	2721	-14.8	*
15.	83	26	7	3612	3672	-29.4	*
16.	93	13	26	3968	4020	-20.0	

Table 10. List of direct repeats in the LaBelle mitochondrial plasmid.

	Correct matches in %	# of correct matches out of total length	Start & stop position of repeat units		Regions shown in Figs.3&7
			first	second	
1.	88	15/17	817-833	1150-1166	
2.	88	15/17	1454-1469	2831-2847	
3.	88	15/17	1514-1530	3781-3795	
4.	84	16/19	142-160	3634-3652	IV
5.	84	16/19	390-407	445-463	
6.	84	16/19	364-382	1390-1406	
7.	84	16/19	3207-3224	3536-3554	
8.	84	21/25	3481-3505	4032-4055	
9.	83	15/18	7-24	1344-1361	
10.	83	15/18	238-255	3264-3279	
11.	83	15/18	461-478	1470-1486	
12.	83	15/18	307-324	2899-2914	
13.	83	15/18	354-371	3745-3759	I
14.	83	15/18	658-672	1424-1441	
15.	83	15/18	676-693	3386-3400	
16.	83	15/18	2155-2172	2988-3005	
17.	83	15/18	2177-2194	4020-4034	
18.	82	18/22	3572-3592	3906-3927	V
19.	81	17/21	106-125	888-906	
20.	81	17/21	185-205	3402-3422	
21.	81	21/26	155-180	3571-3595	
22.	81	21/26	387-412	1138-1162	
23.	80	16/20	234-253	949-966	
24.	80	16/20	385-404	1389-1406	
25.	80	16/20	392-411	3491-3510	
26.	80	16/20	395-414	953-971	
27.	80	16/20	443-462	3403-3421	
28.	80	16/20	556-575	1214-1232	
29.	80	16/20	564-583	2004-2020	
30.	80	16/20	909-928	1164-1181	
31.	80	16/20	970-989	3039-3057	
32.	80	16/20	1207-1225	3269-3286	
33.	80	16/20	1271-1287	1743-1762	
34.	80	16/20	1747-1765	1894-1913	III
35.	80	16/20	1899-1917	3265-3284	
36.	80	16/20	2009-2028	3683-3702	
37.	80	16/20	2272-2291	3082-3101	
38.	80	16/20	2570-2589	3398-3415	
39.	79	19/24	1035-1057	1157-1180	
40.	78	18/23	433-454	1334-1356	
41.	78	18/23	445-465	1275-1295	
42.	78	18/23	696-715	1229-1251	II
43.	78	18/23	1218-1239	3344-3366	
44.	78	18/23	1817-1839	2288-2309	
45.	77	17/22	393-411	1710-1731	
46.	77	17/22	2234-2255	2894-2915	
47.	76	19/25	120-143	2331-2355	
48.	76	19/25	1973-1997	3392-3414	

Table 11. List of regions in the 1,151 amino acid polypeptide of the LaBelle mitochondrial plasmid that are homologous to regions of various RNA polymerases. The percentage identity and number of correct matches in the regions under consideration are shown.

	Amino acids of 1,151 amino acid polypeptide	Amino acids of various		Identity in %	Region in Fig.9	Match length	Colinear regions
		polymerases	regions				
1.	3-51	rpoC	938-988	23.2	a	13/56	*
2.	50-108	rpoC	1051-1109	36.1	d	22/61	*
3.	63-119	rpoC	1061-1123	35.9	e	23/64	
4.	195-236	rpoC	1295-1336	28.9	f	13/45	*
5.	309-357	rpoC	377-429	21.8	h	12/55	
6.	341-374	rpoC	188-222	31.4	i	11/35	
7.	358-402	rpoC	838-885	16.7	l	8/48	
8.	535-582	rpoC	1096-1143	22.0	n	11/50	
9.	854-898	rpoC	1086-1132	30.0	p	15/50	
10.	18-63	mt pol	1-46	34.8	b	16/46	*
11.	27-69	mt pol	737-779	27.9	c	12/43	
12.	254-300	mt pol	328-374	25.0	g	12/48	*
13.	342-388	mt pol	744-793	20.0	k	10/50	
14.	418-461	mt pol	474-517	26.1	m	12/46	*
15.	794-836	mt pol	611-654	28.9	o	13/45	*
16.	857-904	mt pol	736-783	25.0	q	12/48	*
17.	1004-1050	mt pol	859-907	21.6	s	11/51	*
18.	1010-1067	mt pol	832-892	23.0	t	14/61	
19.	341-384	RF862	405-447	30.4	j	14/46	*
20.	943-989	RF862	600-635	23.4	r	11/47	*
21.	469-502	T3 pol	843-878	27.0		10/37	
22.	573-596	T3 pol	779-804	25.9		7/27	
23.	1078-1121	T3 pol	571-614	28.9		13/45	
24.	54-72	T7 pol	704-723	35.0		7/20	
25.	113-149	T7 pol	53-89	28.9		11/38	
26.	572-611	T7 pol	73-112	28.6		12/42	
27.	486-527	SP6 pol	57-104	22.9		11/48	
	rpoC	mt pol					
1.	278-324	"	1237-1283	29.2	u	14/48	
2.	455-502	"	636-681	29.4	v	15/51	*
3.	861-909	"	1202-1251	22.6	w	12/53	*
4.	982-1029	"	465-512	30.6	x	12/53	

rpoC	1407 amino acid long *E. coli* RNA polymerase β' subunit.
mt pol	1351 amino acid long yeast mitochondrial RNA polymerase.
T7 pol	883 amino acid long bacteriophage T7 RNA polymerase.
T3 pol	884 amino acid long bacteriophage T3 RNA polymerase.
SP6 pol	874 amino acid long bacteriophage SP6 RNA polymerase.
RF862	862 amino acid long reading frame between two stop codons in *N. tabacum* chloroplast which is part of a 1360 amino acid long region that corresponds to rpoC gene.

Table 12. Insertions present in the *N. intermedia* EcoRI-6 mtDNA fragment as compared to the homologous region of the LaBelle mitochondrial plasmid (or vice versa).

Insertion				Gap begins at nucleotide number
N. intermedia EcoRI-6 mtDNA fragment				LaBelle mt plasmid
Position	Length in base pairs	GC content in %	Palindromic sequence	
244-276	33	91	present	3092
493-525	33	64	absent	3300
567-594	28	89	present	3341
805-814	10	30	absent	3552
865-937	73	32	absent	3601
1088-1163	76	12	absent	3757
LaBelle mt plasmid				*N. intermedia* EcoRI-6 mtDNA fragment
3902-3927	26	46	absent	1312

Table 13. List of palindromic sequences found in the region of homology between the *N. intermedia* EcoRI-6 mtDNA fragment and the LaBelle mitochondrial plasmid.

N. intermedia EcoRI-6 mtDNA fragment		LaBelle mitochondrial plasmid	
Nucleotide position	Length of stem loop in base pairs	Nucleotide position	Length of stem loop in base pairs
244-276	33	--	--
376-392	17	--	--
467-512	46	--	--
569-596	28	--	--
947-975, 983-1010	29 28	3612-3672	61
1137-1162	26	--	--

-- no corresponding palindromic region

FIGURES

Figure 1.

Sequencing strategy used for the LaBelle mitochondrial plasmid

The plasmid is represented in a linear form.
Only the restriction sites used for cloning are shown in this figure.

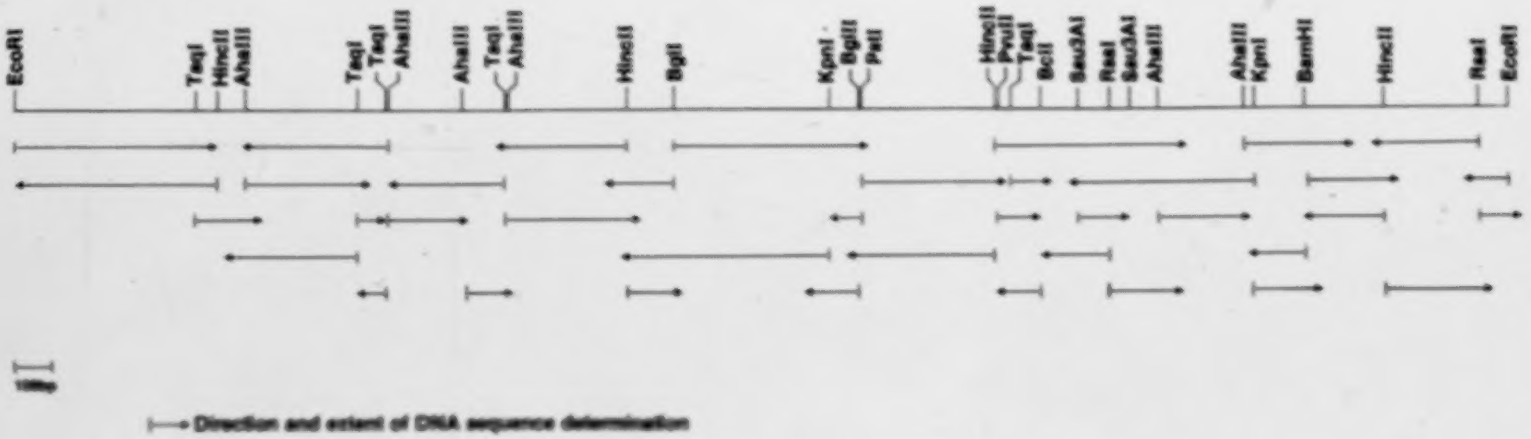

EcoRI
TaqI
HincII
AhaIII
TaqI
TaqI
AhaIII
AhaIII
TaqI
AhaIII
HincII
BglI
KpnI
BglII
PstI
HincII
PvuII
TaqI
BclI
Sau3AI
RsaI
Sau3AI
AhaIII
AhaIII
KpnI
BamHI
HincII
RsaI
EcoRI
100bp
Direction and extent of DNA sequence determination

Figure 2.

DNA sequence of the LaBelle mitochondrial plasmid.

The unique EcoRI site is arbitrarily designated as position 1. The numbers on top of every 10th residue refer to the nucleotide position from this EcoRI site in a clockwise direction.

```
        10         20         30         40         50         60
GAATTCAGTG GGATATGACT CCAGACAGGT CCGGGAGCCA AGTCTAGTGA AACCTCCGAT
CTTAAGTCAC CCTATACTGA GGTCTGTCCA GGCCCTCGGT TCAGATCACT TTGGAGGCTA

        70         80         90        100        110        120
GTGTCGGCCG AAGAAAACTA GAAATCGGGG CCATAGTAAT TCTAAATTCA TGATTAAACA
CACAGCCGGC TTCTTTTGAT CTTTAGCCCC GGTATCATTA AGATTTAAGT ACTAATTTGT

       130        140        150        160        170        180
AAGTTTTAAA GACTTAAATA AAAAAAGTAT ACTCTTTTTG AATAGTGTAC TCGTTAAAAA
TTCAAAATTT CTGAATTTAT TTTTTTCATA TGAGAAAAAC TTATCACATG AGCAATTTTT

       190        200        210        220        230        240
TAGCGGTTTA TTAAAAATAT CACCAGATTT CGAGATAAAA GGTCTTGAAC ATGTCAAAAA
ATCGCCAAAT AATTTTTATA GTGGTCTAAA GCTCTATTTT CCAGAACTTG TACAGTTTTT

       250        260        270        280        290        300
ACTTGTAAGT GATATTGAAA TTTTGGTAAA GTGTTTACCG ATGCTTAGTG ATACGGGGTA
TGAACATTCA CTATAACTTT AAAACCATTT CACAAATGGC TACGAATCAC TATGCCCCAT

       310        320        330        340        350        360
TATTAATAAA AACGATTTTA ATATACCTAT GGTTGATTAT GTAGCCCAGG TATTTTTGGA
ATAATTATTT TTGCTAAAAT TATATGGATA CCAACTAATA CATCGGGTCC ATAAAAACCT

       370        380        390        400        410        420
TAAAAAAGAT AAAGTTCAGT TATGTAAAAT GGATAAAATA GATACTGATA AACGAAAAAA
ATTTTTTCTA TTTCAAGTCA ATACATTTTA CCTATTTTAT CTATGACTAT TTGCTTTTTT

       430        440        450        460        470        480
TTATTCAGTA GCTTTTAATA AAGTTGTATA TAATATGATA CTGTTGTTTC TTTTGTATCT
AATAAGTCAT CGAAAATTAT TTCAACATAT ATTATACTAT GACAACAAAG AAAACATAGA

       490        500        510        520        530        540
AATTACCAGG ATCGAGTTGT GGTTTGTAGA TTGTGATAAG CGTGATGATA TGTTATTGGA
TTAATGGTCC TAGCTCAACA CCAAACATCT AACACTATTC GCACTACTAT ACAATAACCT

       550        560        570        580        590        600
TGTCGCCGTT GACGTTAAAA ACTTAAAGAG GTCTATCACT CTTATGATAC ATCCATACAT
ACAGCGGCAA CTGCAATTTT TGAATTTCTC CAGATAGTGA GAATACTATG TAGGTATGTA

       610        620        630        640        650        660
AAAAAGTAAG CTGGGGAACC GTGGTCTCTC TATAATTTTA AACGGGTTTT CAGGTTACGA
TTTTTCATTC GACCCCTTGG CACCAGAGAG ATATTAAAAT TTGCCCAAAA GTCCAATGCT

       670        680        690        700        710        720
TACTGAATAT TACACACTAG ACGAAACCAA GAAACTGAAT GAGTTGTTAA GTGGTCAAAT
ATGACTTATA ATGTGTGATC TGCTTTGGTT CTTTGACTTA CTCAACAATT CACCAGTTTA

       730        740        750        760        770        780
AGCTGCTAAT ACAGCTATTT ACCTAAGAGT GCCAGTGGTT AATGCAATTC CACTTCATTA
TCGACGATTA TGTCGATAAA TGGATTCTCA CGGTCACCAA TTACGTTAAG GTGAAGTAAT

       790        800        810        820        830        840
TACAGACCTT AGCAGAGATG ATTGTTTAAT GTGGGGTGAT ACTGAAAAAT TGAAAACTGC
ATGTCTGGAA TCGTCTCTAC TAACAAATTA CACCCCACTA TGACTTTTTA ACTTTTGACG

       850        860        870        880        890        900
TTTGCCCAGT ATGGATATAG TCATAAGCAG TATAAGAGCT TTGTTGTATT CAGATCAAGA
AAACGGGTCA TACCTATATC AGTATTCGTC ATATTCTCGA AACAACATAA GTCTAGTTCT
```

```
       910        920        930        940        950        960
CAAGTTGTTG AAAGATATAG TATCAAAATT CGAAGTTATA CCCAATGTTC AAAAAATTTA
GTTCAACAAC TTTCTATATC ATAGTTTTAA GCTTCAATAT GGGTTACAAG TTTTTTAAAT

       970        980        990       1000       1010       1020
TATGATTAAC GGTGATGTGG TATTCGCTTT CCCTAAAACA GAACGTCGAG AGTTGTTTAA
ATACTAATTG CCACTACACC ATAAGCGAAA GGGATTTTGT CTTGCAGCTC TCAACAAATT

      1030       1040       1050       1060       1070       1080
ATTGTTTCAA CCTGGAGAAA AATACAGTAG TATTGAATTG ATGAATGTTT GTGATAATCT
TAACAAAGTT GGACCTCTTT TTATGTCATC ATAACTTAAC TACTTACAAA CACTATTAGA

      1090       1100       1110       1120       1130       1140
AGTTGATAAC GATCTTGAGG TATCCTTGAC TATTCTGATG GCGTTGTTGA GAGAAGTAAG
TCAACTATTG CTAGAACTCC ATAGGAACTG ATAAGACTAC CGCAACAACT CTCTTCATTC

      1150       1160       1170       1180       1190       1200
TGGCAAAGAT GATACTGATA AAATGACAGA TAGTATTAAA AGATGCATCA ACCAGCCTTC
ACCGTTTCTA CTATGACTAT TTTACTGTCT ATCATAATTT TCTACGTAGT TGGTCGGAAG

      1210       1220       1230       1240       1250       1260
GTCTCGTATA ATATATAAAT TTAAAGGGTG TAATGATGTT GATAAAGTGT TGAGTATTTC
CAGAGCATAT TATATATTTA AATTTCCCAC ATTACTACAA CTATTTCACA ACTCATAAAG

      1270       1280       1290       1300       1310       1320
GGTTAAACGT GTGATGTATA TAATGATGCA CTTTTCACCA GCTGAACTAA GCATGTTAAG
CCAATTTGCA CACTACATAT ATTACTACGT GAAAAGTGGT CGACTTGATT CGTACAATTC

      1330       1340       1350       1360       1370       1380
CGATTTCGAA AGTTTTAAAG AAGAGTTGGA TATGATTCAA GGGCAGTATG TGACGCGTGG
GCTAAAGCTT TCAAAATTTC TTCTCAACCT ATACTAAGTT CCCGTCATAC ACTGCGCACC

      1390       1400       1410       1420       1430       1440
TAAACCTCTA AAATATAAAG TAGTTAAGAG TCATGTGCAT ATTCGTGATA CTGAATTATT
ATTTGGAGAT TTTATATTTC ATCAATTCTC AGTACACGTA TAAGCACTAT GACTTAATAA

      1450       1460       1470       1480       1490       1500
ATCTCCTCCT GGTAGCAAAT CTTTAGCAGC TGTTGGTTCT ATTTATGCTG ATGAAGGTTA
TAGAGGAGGA CCATCGTTTA GAAATCGTCG ACAACCAAGA TAAATACGAC TACTTCCAAT

      1510       1520       1530       1540       1550       1560
TAAAAAGGTG GATATTAAGG AGTATAGAGG TAAAATGAAA GAGTTGCTGA ATGAAAAGCC
ATTTTTCCAC CTATAATTCC TCATATCTCC ATTTTACTTT CTCAACGACT TACTTTTCGG

      1570       1580       1590       1600       1610       1620
CAAATTGTTT AAGGAATACG GGATGATGGA TGCTTATATT ACATTAAAAC ATGGAAATAC
GTTTAACAAA TTCCTTATGC CCTACTACCT ACGAATATAA TGTAATTTTG TACCTTTATG

      1630       1640       1650       1660       1670       1680
AATGGAAGAG TTTTATTTAA GCCTTAGCAA GTTAGGAGTG CCGTTAACCT CTACTGGGAT
TTACCTTCTC AAAATAAATT CGGAATCGTT CAATCCTCAC GGCAATTGGA GATGACCCTA

      1690       1700       1710       1720       1730       1740
TAGTAAGTCT TATGTATTCC TAGAATGACA TAAAACTAGA TATGCTGGTT ATCACCTAAT
ATCATTCAGA ATACATAAGG ATCTTACTGT ATTTTGATCT ATACGACCAA TAGTGGATTA

      1750       1760       1770       1780       1790       1800
TGGTGATGTA AACATAAGTG ATTTGGCTTC ATTAATAACA CCTAAAAAAG CCAGGGCGGC
ACCACTACAT TTGTATTCAC TAAACCGAAG TAATTATTGT GGATTTTTTC GGTCCCGCCG
```

```
      1810       1820       1830       1840       1850       1860
TGATGTGTCT AATTATATAA TGTCGTATGT GTCAGCATAC CGAGGAGGTA GAAACGAATC
ACTACACAGA TTAATATATT ACAGCATACA CAGTCGTATG GCTCCTCCAT CTTTGCTTAG

      1870       1880       1890       1900       1910       1920
ATTTATGTAT GGTATTGACG ATATACCTGA GTGTGAACAA CGTAAATGAT TTGATTATGA
TAAATACATA CCATAACTGC TATATGGACT CACACTTGTT GCATTTACTA AACTAATACT

      1930       1940       1950       1960       1970       1980
TATGGTTAGT GCTTACACGA GTGTGATGTC TCTATTGGGT CATCCTGATA TGGAAAAAGC
ATACCAATCA CGAATGTGCT CACACTACAG AGATAACCCA GTAGGACTAT ACCTTTTTCG

      1990       2000       2010       2020       2030       2040
AGGTCGTTTA TATGATAAGA AAGTAAAGAC TATGACTCTT GAAGACATAA TGTTAAATTA
TCCAGCAAAT ATACTATTCT TTCATTTCTG ATACTGAGAA CTTCTGTATT ACAATTTAAT

      2050       2060       2070       2080       2090       2100
TGTCGTTTTA GATGTAGTGT TTAAGTTCCC TAATGGAACT AAGTATCCAT GTATCCCAAC
ACAGCAAAAT CTACATCACA AATTCAAGGG ATTACCTTGA TTCATAGGTA CATAGGGTTG

      2110       2120       2130       2140       2150       2160
TAGAGTTGAT GATAATGTGG ATATTTACCC TTTAGAGGGT CGTAGTATAA TAACAGGGTG
ATCTCAACTA CTATTACACC TATAAATGGG AAATCTCCCA GCATCATATT ATTGTCCCAC

      2170       2180       2190       2200       2210       2220
TGAGTATTGA GTTGCTAAGT CTATGGGTTG TGAAATGACT GTTAAGAGTG GGGTGATGGT
ACTCATAACT CAACGATTCA GATACCCAAC ACTTTACTGA CAATTCTCAC CCCACTACCA

      2230       2240       2250       2260       2270       2280
ACCGTTTGTC AAAGATCGTA AAAGAATTGA AAAAGATCGT CTGGATTATA AAAGTCCCTA
TGGCAAACAG TTTCTAGCAT TTTCTTAACT TTTTCTAGCA GACCTAATAT TTTCAGGGAT

      2290       2300       2310       2320       2330       2340
TAGACAAATA ATGTCAGATC TGCAGCGTAA AAGACGTTCT CATCCAAAGA AAAGTTTTTA
ATCTGTTTAT TACAGTCTAG ACGTCGCATT TTCTGCAAGA GTAGGTTTCT TTTCAAAAAT

      2350       2360       2370       2380       2390       2400
TGATCTTATG TATAAGTTGA TTGGTAATTC TATATACGGT TTAGCTTCTC AGGGTATTGG
ACTAGAATAC ATATTCAACT AACCATTAAG ATATATGCCA AATCGAAGAG TCCCATAACC

      2410       2420       2430       2440       2450       2460
TGGTAAAAAG ACTTTTGATA TAAAAACAAA ATCGTATGTC AAAGTGAGTG GGGGTGAGTT
ACCATTTTTC TGAAAACTAT ATTTTTGTTT TAGCATACAG TTTCACTCAC CCCCACTCAA

      2470       2480       2490       2500       2510       2520
ATCCAATCCA ATATTGGCCA GTTACATAAC TGGTTATTGC AGGGCTTTAA TGGGTGAGTG
TAGGTTAGGT TATAACCGGT CAATGTATTG ACCAATAACG TCCCGAAATT ACCCACTCAC

      2530       2540       2550       2560       2570       2580
TCTTAACAAT GTAGAGAAGT TGGGTGGTAA AGCAATTTCT GTAACAACCG ACGGGTTTTT
AGAATTGTTA CATCTCTTCA ACCCACCATT TCGTTAAAGA CATTGTTGGC TGCCCAAAAA

      2590       2600       2610       2620       2630       2640
AACTAATATT GAAGGTTTTG AGGATAAAAT ATTAAAAAAT GACTCCTTAA ATAAAGATTG
TTGATTATAA CTTCCAAAAC TCCTATTTTA TAATTTTTTA CTGAGGAATT TATTTCTAAC

      2650       2660       2670       2680       2690       2700
TTTGATGTTG TACCGAGATA TTAGAAATCG GTTAACAGCT GTTTTAGACG ACAAAGGAGC
AAACTACAAC ATGGCTCTAT AATCTTTAGC CAATTGTCGA CAAAATCTGC TGTTTCCTCG
```

```
      2710       2720       2730       2740       2750       2760
GGTCATCGAA TCAGATGACC GTGCTTTGGA GATAAAAAAG GAGGACGGTA AGGGTCTATT
CCAGTAGCTT AGTCTACTGG CACGAAACCT CTATTTTTTC CTCCTGCCAT TCCCAGATAA

      2770       2780       2790       2800       2810       2820
GTCTTGAAAA ACCAGAGGAC AGTTGGGTCT TGATCACACC GTTGAGTTAA AGGCAATGTC
CAGAACTTTT TGGTCTCCTG TCAACCCAGA ACTAGTGTGG CAACTCAATT TCCGTTACAG

      2830       2840       2850       2860       2870       2880
AGGGTTTCAA AGCAAATATT TAGACAGACA ATTCTTAGTT GAAGAGCTTA GGCGTAGGAT
TCCCAAAGTT TCGTTTATAA ATCTGTCTGT TAAGAATCAA CTTCTCGAAT CCGCATCCTA

      2890       2900       2910       2920       2930       2940
CACCGGTCCG GCTGATCTTA AAAAATTTGA ATATGTTGAG TCTCGTTTAC GTAAAGCTAA
GTGGCCAGGC CGACTAGAAT TTTTTAAACT TATACAACTC AGAGCAAATG CATTTCGATT

      2950       2960       2970       2980       2990       3000
GGATATTTAC GATAAAGGGG GTCATGTTCT GATTGAGTAC AGCGACAAGA GTTTTAGTAT
CCTATAAATG CTATTTCCCC CAGTACAAGA CTAACTCATG TCGCTGTTCT CAAAATCATA

      3010       3020       3030       3040       3050       3060
TGAGTATGAT AATAAACGTC GTATTGTCAA TGCTTGATCG GTGTGTGGTA CTCTGTTAGA
ACTCATACTA TTATTTGCAG CATAACAGTT ACGAACTAGC CACACACCAT GAGACAATCT

      3070       3080       3090       3100       3110       3120
GTCTGTACCT TGAGGTTCGA TAACTCCTTA TAGAAGAATA AGGAATTTAA AGGAAACCGT
CAGACATGGA ACTCCAAGCT ATTGAGGAAT ATCTTCTTAT TCCTTAAATT TCCTTTGGCA

      3130       3140       3150       3160       3170       3180
TAATACTCCA GTATTCGTAA AAGGAGCCTT TACTGGTGGG AGTGGAAAAA AGTATAAAAG
ATTATGAGGT CATAAGCATT TTCCTCGGAA ATGACCACCC TCACCTTTTT TCATATTTTC

      3190       3200       3210       3220       3230       3240
TACTATTGAA ACTAGTGTTA GAGGTTTCAT TAAAGCTGCT CTATCTGATA ATGACCGTTT
ATGATAACTT TGATCACAAT CTCCAAAGTA ATTTCGACGA GATAGACTAT TACTGGCAAA

      3250       3260       3270       3280       3290       3300
TGGTATTCCC AAAGATCAAT TTCAAACTTA TAATGATATC ATTAAATTCA TACGCGGTCA
ACCATAAGGG TTTCTAGTTA AAGTTTGAAT ATTACTATAG TAATTTAAGT ATGCGCCAGT

      3310       3320       3330       3340       3350       3360
TGATGCAGCC AAGAATGTTA AACTCTCAAA GACTAGTATT TCCAATTTAA AATGTCGTAA
ACTACGTCGG TTCTTACAAT TTGAGAGTTT CTGATCATAA AGGTTAAATT TTACAGCATT

      3370       3380       3390       3400       3410       3420
TGCTATAGCA CGAACGGTAC CTCGTACAGA CGAAACAGAA GGGTTTATAA AATATGTCAC
ACGATATCGT GCTTGCCATG GAGCATGTCT GCTTTGTCTT CCCAAATATT TTATACAGTG

      3430       3440       3450       3460       3470       3480
TAAAAAAATA ACCACCTTCG ACGTTGATCG TTTCTTTAAG GAAAAACCTC AAGAAAAAGC
ATTTTTTTAT TGGTGGAAGC TGCAACTAGC AAAGAAATTC CTTTTTGGAG TTCTTTTTCG

      3490       3500       3510       3520       3530       3540
GAACAAAAAA GATAGGATAG AAAGTGATAA TTAGAGGATC CCCCTTTCAA TATTATCTTT
CTTGTTTTTT CTATCCTATC TTTCACTATT AATCTCCTAG GGGGAAAGTT ATAATAGAAA

      3550       3560       3570       3580       3590       3600
AATAGCAGCT CTATTCATTT TTTAAGTATA TTTATTAGTA GTGTATCGTT AGAAAAAAAT
TTATCGTCGA GATAAGTAAA AAATTCATAT AAATAATCAT CACATAGCAA TCTTTTTTTA
```

```
      3610       3620       3630       3640       3650       3660
CTAAAAAAGG TGTCCACATT TTTTGGAAAC CATAAAAAGT CTACATTTTT TGGTTTCTAG
GATTTTTTCC ACAGGTGTAA AAAACCTTTG GTATTTTTCA GATGTAAAAA ACCAAAGATC

      3670       3680       3690       3700       3710       3720
AAAAATGTGG ACGGTAGATA TAACTATAAC GTTTGTAGAC ATGCGGGTGG GTGGGGACCA
TTTTTACACC TGCCATCTAT ATTGATATTG CAAACATCTG TACGCCCACC CACCCCTGGT

      3730       3740       3750       3760       3770       3780
TGACATTACG TCAGGTTAAC GAAGTTTTAT AAAAAAATAC TTCGGGAATT TTGATTATCA
ACTGTAATGC AGTCCAATTG CTTCAAAATA TTTTTTTATG AAGCCCTTAA AACTAATAGT

      3790       3800       3810       3820       3830       3840
ATTAAGAGTA TGAGGGATTC ATGTGTTAAT TAGTGTAAGA AATATTTCTA TTTGATAATT
TAATTCTCAT ACTCCCTAAG TACACAATTA ATCACATTCT TTATAAAGAT AAACTATTAA

      3850       3860       3870       3880       3890       3900
AGATAAGCAT GTTTACAACC AGTGTGGCGT CTCACAATTG TGTTGTTGTT GTTATTGTAA
TCTATTCGTA CAAATGTTGG TCACACCGCA GAGTGTTAAC ACAACAACAA CAATAACATT

      3910       3920       3930       3940       3950       3960
TCGCATTAGT AGTAGTAGTA GCGGTAGTAG TTTGTTTGTT TATGAGCACC CAGCGCTGAC
AGCGTAATCA TCATCATCAT CGCCATCATC AAACAAACAA ATACTCGTGG GTCGCGACTG

      3970       3980       3990       4000       4010       4020
TGCTACCTGG GCTACTTAAA TACCGCTTGC AGTACTGCTG CGGGCTTTTA AGCTAGCCCA
ACGATGGACC CGATGAATTT ATGGCGAACG TCATGACGAC GCCCGAAAAT TCGATCGGGT

      4030       4040       4050       4060       4070
ATCTGGGTTG TGAACAAAAT AGTAGGATGG AGAGTAGGGG AAGATCCAAA
TAGACCCAAC ACTTGTTTTA TCATCCTACC TCTCATCCCC TTCTAGGTTT
```

Figure 3.

Circular restriction map of the LaBelle mitochondrial plasmid and summary of major genetic features.

The 1,151 amino acid ORF is indicated inside the circle with the potential start and stop codons indicated.
The region of the mitochondrial plasmid containing homology to the *Neurospora* mtDNA EcoRI-6 fragment is indicated by the arc outside the circle. Numbers shown correspond to the numbering in Fig. 14 for the region of the *N. intermedia* EcoRI-6 fragment whose sequence was determined.
Restriction site code: B-BamHI; E-EcoRI; H-HincII; and P-PstI.
The stem loop structures indicate regions with potential to form secondary structures that have ΔG equal to or less than -10 Kcal/mol (see text).
The regions marked by "wiggly" lines and labelled with lower case letters are regions that are 60% or more GC rich (see text).
The short straight lines outside the circle indicate direct repeat units where at least one member of the two repeats is associated with a potential hairpin loop forming region. These direct repeat units are labelled with Roman numerals.

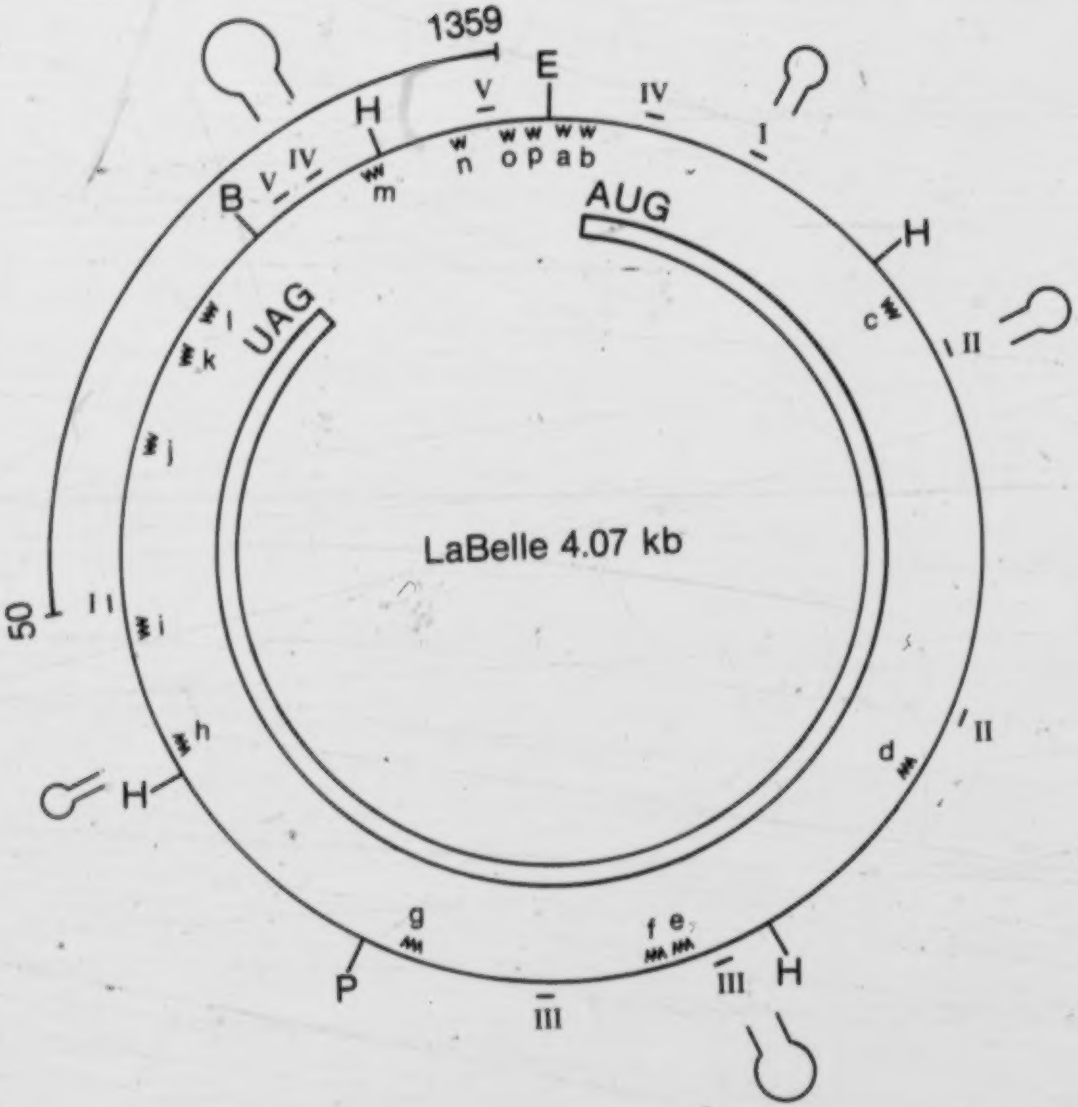
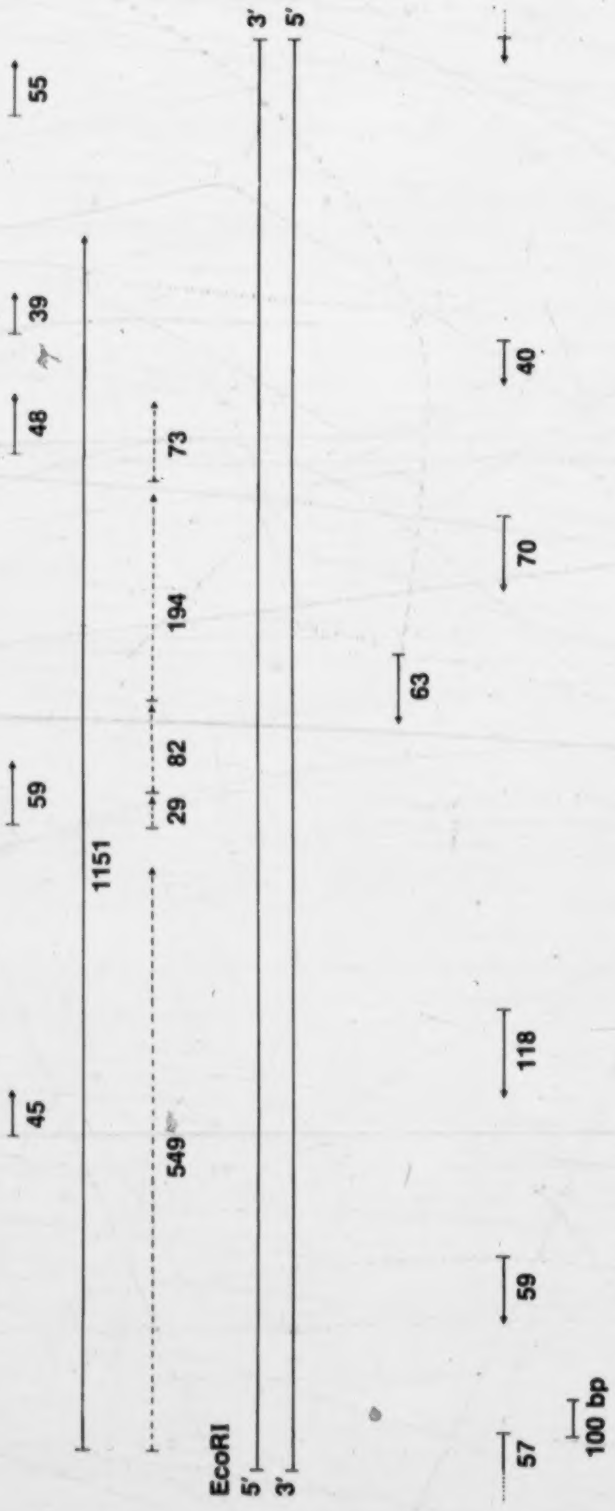

1359
E
V
IV
I
H
B
V
IV
H
w w w w w
n o p a b
m
AUG
UAG
H
c
II
l
k
j
LaBelle 4.07 kb
50
II
i
h
H
II
d
g
f e
P
III
III
H

Figure 4.

Positions and lengths of open reading frames (ORFs) in the LaBelle mitochondrial plasmid.

Broken arrows indicate ORFs arising if the standard genetic code is used to translate the sequence.
Solid arrows indicate ORFs arising if the *Neurospora* mitochondrial genetic code is used to translate the sequences.
Numbers indicate length of ORF in amino acids.


EcoRI
100 bp

- Figure 5.

Amino acid sequence of the polypeptide potentially encoded by the longest open reading frame using the Neurospora mitochondrial genetic code.

The numbers above the base sequence correspond to the numbering of nucleotides as in Fig. 2.
The numbers on the sides refer to the amino acid residues assuming that the met codon at nucleotide position 59-61 is the first amino acid of the 1,151 amino acid long ORF.

```
                                         88                                     118
    ATG TGT CGG CCG AAG AAA ACT AGA AAT CGG GGC CAT AGT AAT TCT AAA TTC ATG ATT AAA
  1 Met Cys Arg Pro Lys Lys Thr Arg Asn Arg Gly His Ser Asn Ser Lys Phe Met Ile Lys  20

                                        148                                     178
    CAA AGT TTT AAA GAC TTA AAT AAA AAA AGT ATA CTC TTT TTG AAT AGT GTA CTC GTT AAA
 21 Gln Ser Phe Lys Asp Leu Asn Lys Lys Ser Ile Leu Phe Leu Asn Ser Val Leu Val Lys  40

                                        208                                     238
    AAT AGC GGT TTA TTA AAA ATA TCA CCA GAT TTC GAG ATA AAA GGT CTT GAA CAT GTC AAA
 41 Asn Ser Gly Leu Leu Lys Ile Ser Pro Asp Phe Glu Ile Lys Gly Leu Glu His Val Lys  60

                                        268                                     298
    AAA CTT GTA AGT GAT ATT GAA ATT TTG GTA AAG TGT TTA CCG ATG CTT AGT GAT ACG GGG
 61 Lys Leu Val Ser Asp Ile Glu Ile Leu Val Lys Cys Leu Pro Met Leu Ser Asp Thr Gly  80

                                        328                                     358
    TAT ATT AAT AAA AAC GAT TTT AAT ATA CCT ATG GTT GAT TAT GTA GCC CAG GTA TTT TTG
 81 Tyr Ile Asn Lys Asn Asp Phe Asn Ile Pro Met Val Asp Tyr Val Ala Gln Val Phe Leu 100

                                        388                                     418
    GAT AAA AAA GAT AAA GTT CAG TTA TGT AAA ATG GAT AAA ATA GAT ACT GAT AAA CGA AAA
101 Asp Lys Lys Asp Lys Val Gln Leu Cys Lys Met Asp Lys Ile Asp Thr Asp Lys Arg Lys 120

                                        448                                     478
    AAT TAT TCA GTA GCT TTT AAT AAA GTT GTA TAT AAT ATG ATA CTG TTG TTT CTT TTG TAT
121 Asn Tyr Ser Val Ala Phe Asn Lys Val Val Tyr Asn Met Ile Leu Leu Phe Leu Leu Tyr 140

                                        508                                     538
    CTA ATT ACC AGG ATC GAG TTG TGG TTT GTA GAT TGT GAT AAG CGT GAT GAT ATG TTA TTG
141 Leu Ile Thr Arg Ile Glu Leu Trp Phe Val Asp Cys Asp Lys Arg Asp Asp Met Leu Leu 160

                                        568                                     598
    GAT GTC GCC GTT GAC GTT AAA AAC TTA AAG AGG TCT ATC ACT CTT ATG ATA CAT CCA TAC
161 Asp Val Ala Val Asp Val Lys Asn Leu Lys Arg Ser Ile Thr Leu Met Ile His Pro Tyr 180

                                        628                                     658
    ATA AAA AGT AAG CTG GGG AAC CGT GGT CTC TCT ATA ATT TTA AAC GGG TTT TCA GGT TAC
181 Ile Lys Ser Lys Leu Gly Asn Arg Gly Leu Ser Ile Ile Leu Asn Gly Phe Ser Gly Tyr 200

                                        688                                     718
    GAT ACT GAA TAT TAC ACA CTA GAC GAA ACC AAG AAA CTG AAT GAG TTG TTA AGT GGT CAA
201 Asp Thr Glu Tyr Tyr Thr Leu Asp Glu Thr Lys Lys Leu Asn Glu Leu Leu Ser Gly Gln 220

                                        748                                     778
    ATA GCT GCT AAT ACA GCT ATT TAC CTA AGA GTG CCA GTG GTT AAT GCA ATT CCA CTT CAT
221 Ile Ala Ala Asn Thr Ala Ile Tyr Leu Arg Val Pro Val Val Asn Ala Ile Pro Leu His 240

                                        808                                     838
    TAT ACA GAC CTT AGC AGA GAT GAT TGT TTA ATG TGG GGT GAT ACT GAA AAA TTG AAA ACT
241 Tyr Thr Asp Leu Ser Arg Asp Asp Cys Leu Met Trp Gly Asp Thr Glu Lys Leu Lys Thr 260

                                        868                                     898
    GCT TTG CCC AGT ATG GAT ATA GTC ATA AGC AGT ATA AGA GCT TTG TTG TAT TCA GAT CAA
261 Ala Leu Pro Ser Met Asp Ile Val Ile Ser Ser Ile Arg Ala Leu Leu Tyr Ser Asp Gln 280

                                        928                                     958
    GAC AAG TTG TTG AAA GAT ATA GTA TCA AAA TTC GAA GTT ATA CCC AAT GTT CAA AAA ATT
281 Asp Lys Leu Leu Lys Asp Ile Val Ser Lys Phe Glu Val Ile Pro Asn Val Gln Lys Ile 300
```

```
                                        988                                    1018
    TAT ATG ATT AAC GGT GAT GTG GTA TTC GCT TTC CCT AAA ACA GAA CGT CGA GAG TTG TTT
301 Tyr Met Ile Asn Gly Asp Val Val Phe Ala Phe Pro Lys Thr Glu Arg Arg Glu Leu Phe  320

                                       1048                                    1078
    AAA TTG TTT CAA CCT GGA GAA AAA TAC AGT AGT ATT GAA TTG ATG AAT GTT TGT GAT AAT
321 Lys Leu Phe Gln Pro Gly Glu Lys Tyr Ser Ser Ile Glu Leu Met Asn Val Cys Asp Asn  340

                                       1108                                    1138
    CTA GTT GAT AAC GAT CTT GAG GTA TCC TTG ACT ATT CTG ATG GCG TTG TTG AGA GAA GTA
341 Leu Val Asp Asn Asp Leu Glu Val Ser Leu Thr Ile Leu Met Ala Leu Leu Arg Glu Val  360

                                       1168                                    1198
    AGT GGC AAA GAT GAT ACT GAT AAA ATG ACA GAT AGT ATT AAA AGA TGC ATC AAC CAG CCT
361 Ser Gly Lys Asp Asp Thr Asp Lys Met Thr Asp Ser Ile Lys Arg Cys Ile Asn Gln Pro  380

                                       1228                                    1258
    TCG TCT CGT ATA ATA TAT AAA TTT AAA GGG TGT AAT GAT GTT GAT AAA GTG TTG AGT ATT
381 Ser Ser Arg Ile Ile Tyr Lys Phe Lys Gly Cys Asn Asp Val Asp Lys Val Leu Ser Ile  400

                                       1288                                    1318
    TCG GTT AAA CGT GTG ATG TAT ATA ATG ATG CAC TTT TCA CCA GCT GAA CTA AGC ATG TTA
401 Ser Val Lys Arg Val Met Tyr Ile Met Met His Phe Ser Pro Ala Glu Leu Ser Met Leu  420

                                       1348                                    1378
    AGC GAT TTC GAA AGT TTT AAA GAA GAG TTG GAT ATG ATT CAA GGG CAG TAT GTG ACG CGT
421 Ser Asp Phe Glu Ser Phe Lys Glu Glu Leu Asp Met Ile Gln Gly Gln Tyr Val Thr Arg  440

                                       1408                                    1438
    GGT AAA CCT CTA AAA TAT AAA GTA GTT AAG AGT CAT GTG CAT ATT CGT GAT ACT GAA TTA
441 Gly Lys Pro Leu Lys Tyr Lys Val Val Lys Ser His Val His Ile Arg Asp Thr Glu Leu  460

                                       1468                                    1498
    TTA TCT CCT CCT GGT AGC AAA TCT TTA GCA GCT GTT GGT TCT ATT TAT GCT GAT GAA GGT
461 Leu Ser Pro Pro Gly Ser Lys Ser Leu Ala Ala Val Gly Ser Ile Tyr Ala Asp Glu Gly  480

                                       1528                                    1558
    TAT AAA AAG GTG GAT ATT AAG GAG TAT AGA GGT AAA ATG AAA GAG TTG CTG AAT GAA AAG
481 Tyr Lys Lys Val Asp Ile Lys Glu Tyr Arg Gly Lys Met Lys Glu Leu Leu Asn Glu Lys  500

                                       1588                                    1618
    CCC AAA TTG TTT AAG GAA TAC GGG ATG ATG GAT GCT TAT ATT ACA TTA AAA CAT GGA AAT
501 Pro Lys Leu Phe Lys Glu Tyr Gly Met Met Asp Ala Tyr Ile Thr Leu Lys His Gly Asn  520

                                       1648                                    1678
    ACA ATG GAA GAG TTT TAT TTA AGC CTT AGC AAG TTA GGA GTG CCG TTA ACC TCT ACT GGG
521 Thr Met Glu Glu Phe Tyr Leu Ser Leu Ser Lys Leu Gly Val Pro Leu Thr Ser Thr Gly  540

                                       1708                                    1738
    ATT AGT AAG TCT TAT GTA TTC CTA GAA TGA CAT AAA ACT AGA TAT GCT GGT TAT CAC CTA
541 Ile Ser Lys Ser Tyr Val Phe Leu Glu Trp His Lys Thr Arg Tyr Ala Gly Tyr His Leu  560

                                       1768                                    1798
    ATT GGT GAT GTA AAC ATA AGT GAT TTG GCT TCA TTA ATA ACA CCT AAA AAA GCC AGG GCG
561 Ile Gly Asp Val Asn Ile Ser Asp Leu Ala Ser Leu Ile Thr Pro Lys Lys Ala Arg Ala  580

                                       1828                                    1858
    GCT GAT GTG TCT AAT TAT ATA ATG TCG TAT GTG TCA GCA TAC CGA GGA GGT AGA AAC GAA
581 Ala Asp Val Ser Asn Tyr Ile Met Ser Tyr Val Ser Ala Tyr Arg Gly Gly Arg Asn Glu  600
```

```
                                         1888                                            1918
      TCA TTT ATG TAT GGT ATT GAC GAT ATA CCT GAG TGT GAA CAA CGT AAA TGA TTT GAT TAT
601   Ser Phe Met Tyr Gly Ile Asp Asp Ile Pro Glu Cys Glu Gln Arg Lys Trp Phe Asp Tyr   620

                                         1948                                            1978
      GAT ATG GTT AGT GCT TAC ACG AGT GTG ATG TCT CTA TTG GGT CAT CCT GAT ATG GAA AAA
621   Asp Met Val Ser Ala Tyr Thr Ser Val Met Ser Leu Leu Gly His Pro Asp Met Glu Lys   640

                                         2008                                            2038
      GCA GGT CGT TTA TAT GAT AAG AAA GTA AAG ACT ATG ACT CTT GAA GAC ATA ATG TTA AAT
641   Ala Gly Arg Leu Tyr Asp Lys Lys Val Lys Thr Met Thr Leu Glu Asp Ile Met Leu Asn   660

                                         2068                                            2098
      TAT GTC GTT TTA GAT GTA GTG TTT AAG TTC CCT AAT GGA ACT AAG TAT CCA TGT ATC CCA
661   Tyr Val Val Leu Asp Val Val Phe Lys Phe Pro Asn Gly Thr Lys Tyr Pro Cys Ile Pro   680

                                         2128                                            2158
      ACT AGA GTT GAT GAT AAT GTG GAT ATT TAC CCT TTA GAG GGT CGT AGT ATA ATA ACA GGG
681   Thr Arg Val Asp Asp Asn Val Asp Ile Tyr Pro Leu Glu Gly Arg Ser Ile Ile Thr Gly   700

                                         2188                                            2218
      TGT GAG TAT TGA GTT GCT AAG TCT ATG GGT TGT GAA ATG ACT GTT AAG AGT GGG GTG ATG
701   Cys Glu Tyr Trp Val Ala Lys Ser Met Gly Cys Glu Met Thr Val Lys Ser Gly Val Met   720

                                         2248                                            2278
      GTA CCG TTT GTC AAA GAT CGT AAA AGA ATT GAA AAA GAT CGT CTG GAT TAT AAA AGT CCC
721   Val Pro Phe Val Lys Asp Arg Lys Arg Ile Glu Lys Asp Arg Leu Asp Tyr Lys Ser Pro   740

                                         2308                                            2338
      TAT AGA CAA ATA ATG TCA GAT CTG CAG CGT AAA AGA CGT TCT CAT CCA AAG AAA AGT TTT
741   Tyr Arg Gln Ile Met Ser Asp Leu Gln Arg Lys Arg Arg Ser His Pro Lys Lys Ser Phe   760

                                         2368                                            2398
      TAT GAT CTT ATG TAT AAG TTG ATT GGT AAT TCT ATA TAC GGT TTA GCT TCT CAG GGT ATT
761   Tyr Asp Leu Met Tyr Lys Leu Ile Gly Asn Ser Ile Tyr Gly Leu Ala Ser Gln Gly Ile   780

                                         2428                                            2458
      GGT GGT AAA AAG ACT TTT GAT ATA AAA ACA AAA TCG TAT GTC AAA GTG AGT GGG GGT GAG
781   Gly Gly Lys Lys Thr Phe Asp Ile Lys Thr Lys Ser Tyr Val Lys Val Ser Gly Gly Glu   800

                                         2488                                            2518
      TTA TCC AAT CCA ATA TTG GCC AGT TAC ATA ACT GGT TAT TGC AGG GCT TTA ATG GGT GAG
801   Leu Ser Asn Pro Ile Leu Ala Ser Tyr Ile Thr Gly Tyr Cys Arg Ala Leu Met Gly Glu   820

                                         2548                                            2578
      TGT CTT AAC AAT GTA GAG AAG TTG GGT GGT AAA GCA ATT TCT GTA ACA ACC GAC GGG TTT
821   Cys Leu Asn Asn Val Glu Lys Leu Gly Gly Lys Ala Ile Ser Val Thr Thr Asp Gly Phe   840

                                         2608                                            2638
      TTA ACT AAT ATT GAA GGT TTT GAG GAT AAA ATA TTA AAA AAT GAC TCC TTA AAT AAA GAT
841   Leu Thr Asn Ile Glu Gly Phe Glu Asp Lys Ile Leu Lys Asn Asp Ser Leu Asn Lys Asp   860

                                         2668                                            2698
      TGT TTG ATG TTG TAC CGA GAT ATT AGA AAT CGG TTA ACA GCT GTT TTA GAC GAC AAA GGA
861   Cys Leu Met Leu Tyr Arg Asp Ile Arg Asn Arg Leu Thr Ala Val Leu Asp Asp Lys Gly   880

                                         2728                                            2758
      GCG GTC ATC GAA TCA GAT GAC CGT GCT TTG GAG ATA AAA AAG GAG GAC GGT AAG GGT CTA
881   Ala Val Ile Glu Ser Asp Asp Arg Ala Leu Glu Ile Lys Lys Glu Asp Gly Lys Gly Leu   900
```

```
                                           2788                                    2818
        TTG TCT TGA AAA ACC AGA GGA CAG TTG GGT CTT GAT CAC ACC GTT GAG TTA AAG GCA ATG
   901  Leu Ser Trp Lys Thr Arg Gly Gln Leu Gly Leu Asp His Thr Val Glu Leu Lys Ala Met   920

                                           2848                                    2878
        TCA GGG TTT CAA AGC AAA TAT TTA GAC AGA CAA TTC TTA GTT GAA GAG CTT AGG CGT AGG
   921  Ser Gly Phe Gln Ser Lys Tyr Leu Asp Arg Gln Phe Leu Val Glu Glu Leu Arg Arg Arg   940

                                           2908                                    2938
        ATC ACC GGT CCG GCT GAT CTT AAA AAA TTT GAA TAT GTT GAG TCT CGT TTA CGT AAA GCT
   941  Ile Thr Gly Pro Ala Asp Leu Lys Lys Phe Glu Tyr Val Glu Ser Arg Leu Arg Lys Ala   960

                                           2968                                    2998
        AAG GAT ATT TAC GAT AAA GGG GGT CAT GTT CTG ATT GAG TAC AGC GAC AAG AGT TTT AGT
   961  Lys Asp Ile Tyr Asp Lys Gly Gly His Val Leu Ile Glu Tyr Ser Asp Lys Ser Phe Ser   980

                                           3028                                    3058
        ATT GAG TAT GAT AAT AAA CGT CGT ATT GTC AAT GCT TGA TCG GTG TGT GGT ACT CTG TTA
   981  Ile Glu Tyr Asp Asn Lys Arg Arg Ile Val Asn Ala Trp Ser Val Cys Gly Thr Leu Leu  1000

                                           3088                                    3118
        GAG TCT GTA CCT TGA GGT TCG ATA ACT CCT TAT AGA AGA ATA AGG AAT TTA AAG GAA ACC
  1001  Glu Ser Val Pro Trp Gly Ser Ile Thr Pro Tyr Arg Arg Ile Arg Asn Leu Lys Glu Thr  1020

                                           3148                                    3178
        GTT AAT ACT CCA GTA TTC GTA AAA GGA GCC TTT ACT GGT GGG AGT GGA AAA AAG TAT AAA
  1021  Val Asn Thr Pro Val Phe Val Lys Gly Ala Phe Thr Gly Gly Ser Gly Lys Lys Tyr Lys  1040

                                           3208                                    3238
        AGT ACT ATT GAA ACT AGT GTT AGA GGT TTC ATT AAA GCT GCT CTA TCT GAT AAT GAC CGT
  1041  Ser Thr Ile Glu Thr Ser Val Arg Gly Phe Ile Lys Ala Ala Leu Ser Asp Asn Asp Arg  1060

                                           3268                                    3298
        TTT GGT ATT CCC AAA GAT CAA TTT CAA ACT TAT AAT GAT ATC ATT AAA TTC ATA CGC GGT
  1061  Phe Gly Ile Pro Lys Asp Gln Phe Gln Thr Tyr Asn Asp Ile Ile Lys Phe Ile Arg Gly  1080

                                           3328                                    3358
        CAT GAT GCA GCC AAG AAT GTT AAA CTC TCA AAG ACT AGT ATT TCC AAT TTA AAA TGT CGT
  1081  His Asp Ala Ala Lys Asn Val Lys Leu Ser Lys Thr Ser Ile Ser Asn Leu Lys Cys Arg  1100

                                           3388                                    3418
        AAT GCT ATA GCA CGA ACG GTA CCT CGT ACA GAC GAA ACA GAA GGG TTT ATA AAA TAT GTC
  1101  Asn Ala Ile Ala Arg Thr Val Pro Arg Thr Asp Glu Thr Glu Gly Phe Ile Lys Tyr Val  1120

                                           3448                                    3478
        ACT AAA AAA ATA ACC ACC TTC GAC GTT GAT CGT TTC TTT AAG GAA AAA CCT CAA GAA AAA
  1121  Thr Lys Lys Ile Thr Thr Phe Asp Val Asp Arg Phe Phe Lys Glu Lys Pro Gln Glu Lys  1140

                                           3508
        GCG AAC AAA AAA GAT AGG ATA GAA AGT GAT AAT TAG
  1141  Ala Asn Lys Lys Asp Arg Ile Glu Ser Asp Asn End  1152
```

Figure 6.

Hydropathy plot of the putative 1,151 amino acid polypeptide.

Numbers on the x-axis refer to the amino acid residues of the 1,151 amino acid long polypeptide as described in Fig. 5. Numbers on the y-axis refer to the hydrophilicity value assigned to each amino acid residue by the algorithm of Hopp and Woods, 1981.
Regions where there is a high probability of forming α-helices or β-pleated sheets, as predicted by the algorithm of Garnier et al. (1978), are indicated by coils and zigzag lines respectively.

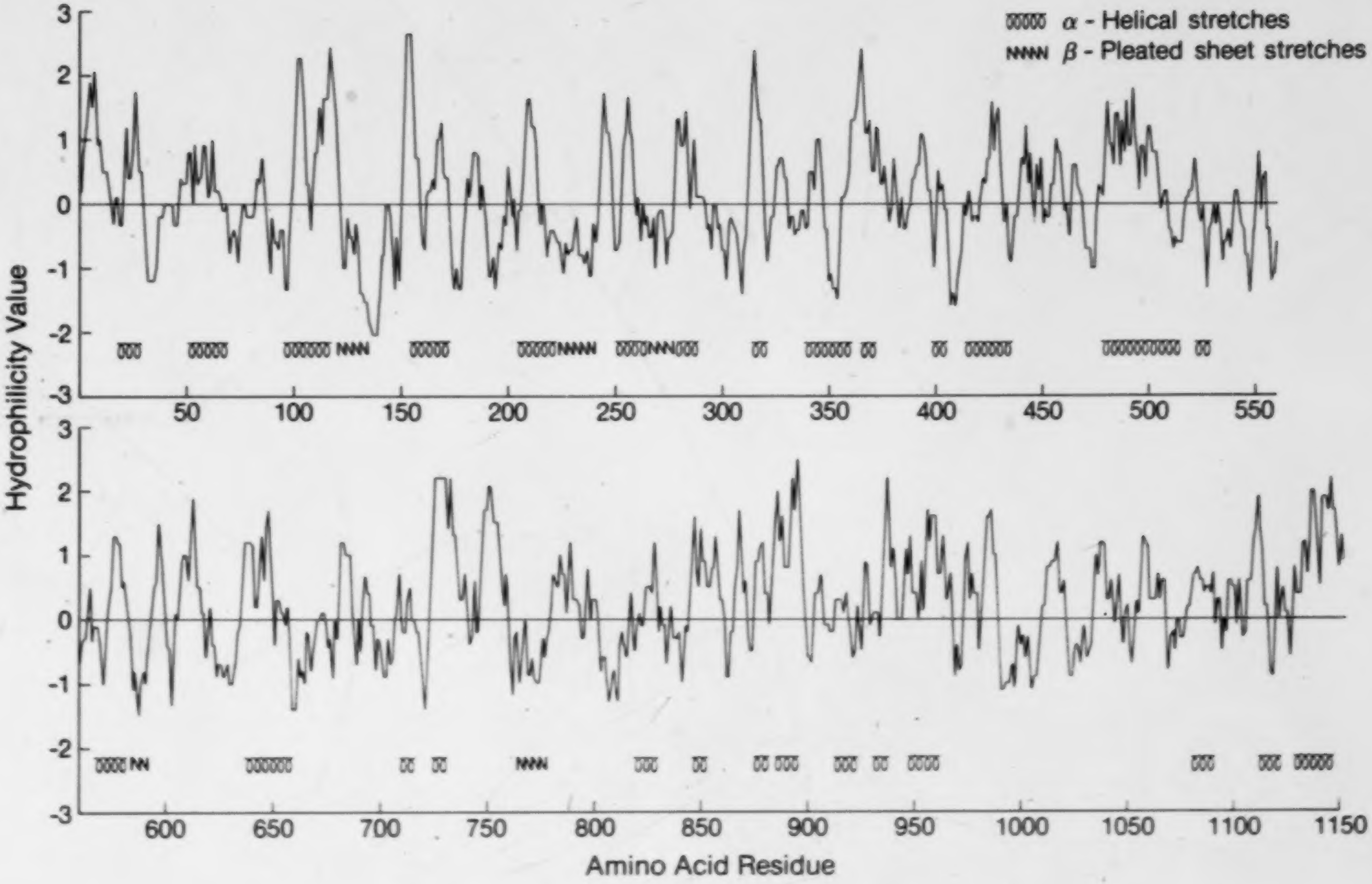

α - Helical stretches
β - Pleated sheet stretches
Hydrophilicity Value
3
2
1
0
-1
-2
-3
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
950
1000
1050
1100
1150
Amino Acid Residue

Figure 7.

Palindromic sequences in the LaBelle mitochondrial plasmid with free energies of formation (ΔG) equal to or less than -10 Kcal/mol.

Numbers immediately below the stem indicate the nucleotide position corresponding to numbering in Fig. 2.
The negative numbers at the bottom are the free energies of formation (ΔG) in Kcal/mol.
The solid lines outside the palindromic sequences show the regions which are part of direct repeats as shown in Fig. 3.
The Roman numerals indicate the pair of direct repeats these regions belong to as illustrated in Fig. 3.

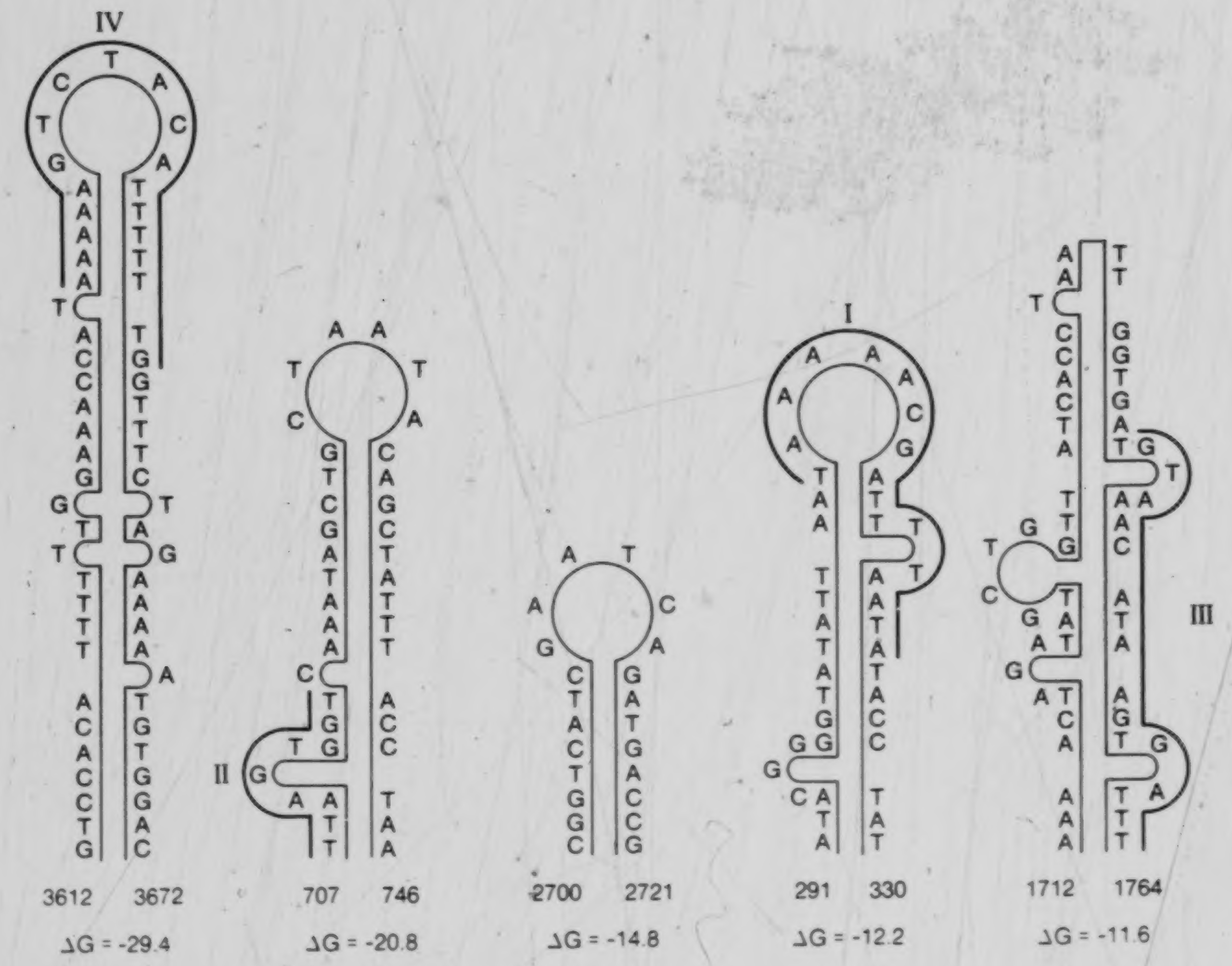

IV
3612
3672
ΔG = -29.4
II
707
746
ΔG = -20.8
2700
2721
ΔG = -14.8
I
291
330
ΔG = -12.2
III
1712
1764
ΔG = -11.6

Figure 8.

Repeated regions and consensus sequences; their similarities to the LaBelle mitochondrial plasmid and the *N. intermedia* EcoRI-6 mtDNA sequences.

The numbers on the extreme left indicate the number of correct matches out of the total.
The value below this number indicates the percentage of correct matches in the given sequence.
The mismatches are indicated by asterisks between the sequences.
Numbers preceeding and following the nucleotide sequences indicate the region under consideration.
The numbering of LaBelle and EcoRI-6 (denoted by Eco6 in this diagram) corresponds to the numbering in Figs. 2 and 15.
The numbering for *Xenopus laevis*, sea urchin and rat sequences are as in the original references (see text).

8a. A direct repeat in the LaBelle plasmid which shows similarities to *X. laevis* and sea urchin sequences.

8b. Similarity between a repeated sequence in the LaBelle plasmid and part of the repetitive 1723 element clone CI7 in *Xenopus laevis* and sea urchin genomic clone Csp 2109B.

8c. Similarity between the LaBelle plasmid, rat cellular DNA adjacent to the right side of an insertion site for SV40, and the EcoRI-6 mtDNA fragment from *N. intermedia* LaBelle-1b strain.

The arrows in 8c indicate the positions where substitutions by the bases shown would restore the sequence to the $(TTG)_6$ repeat.

8d. Consensus sequence in *Torulopsis glabrata* mtDNA and its similarity to the LaBelle plasmid and *N. intermedia* EcoRI-6 mtDNA sequences.

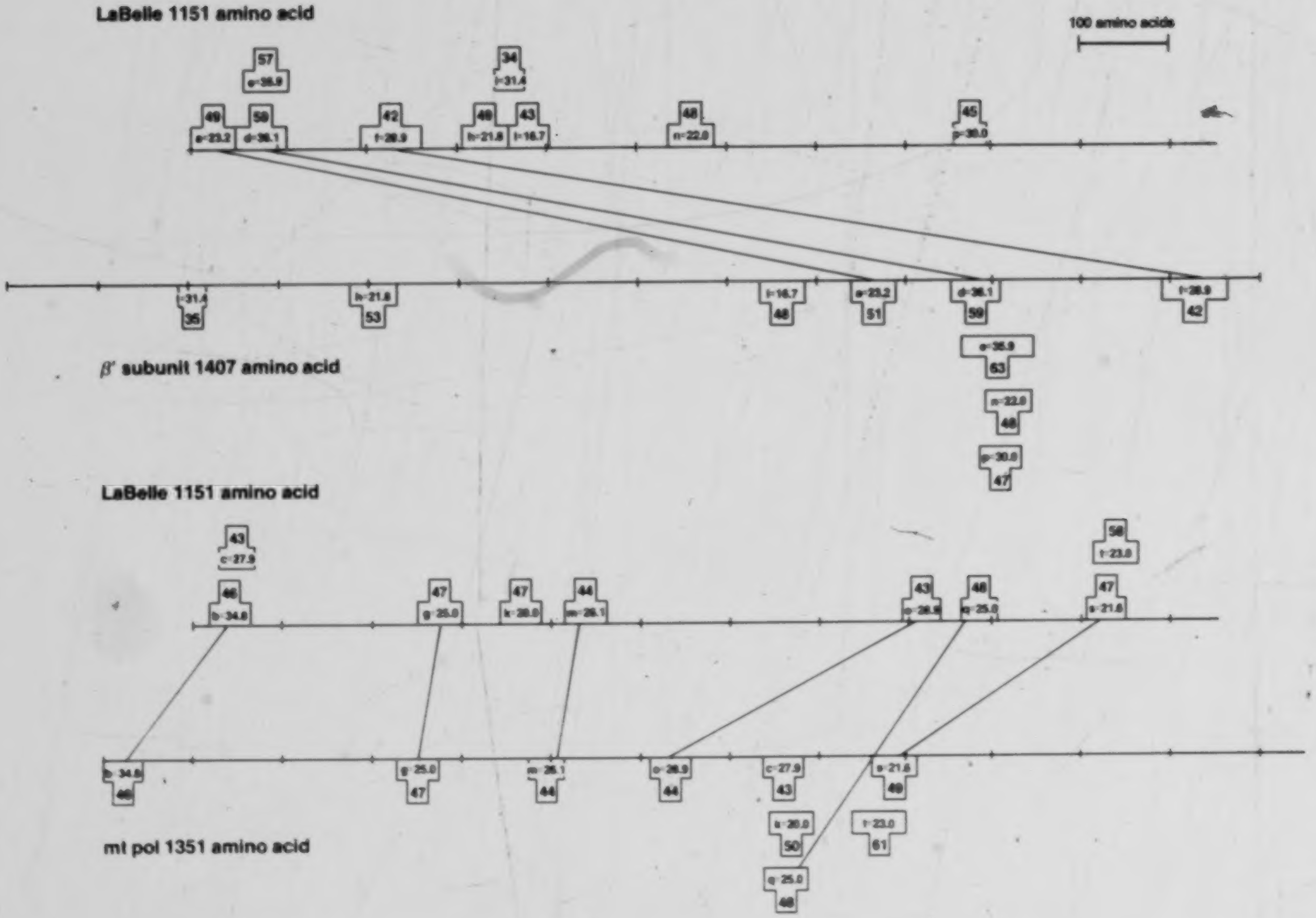
The underlined C is the most conserved nucleotide in this consensus sequence.

```
a)  18/22  LaBelle      3572   TTATTAGTAGT GTATCGTTAG        3592
                                  *       *   *  *
    0.82   LaBelle      3906   TTAGTAGTAGTAGTAGCGGTAG        3927


b)  23/25  LaBelle      3907   TAGTAGTAGTAGTAGCGGTAGTAGT     3931
                                              **
    0.92   X. laevis    542    TAGTAGTAGTAGTAGTAGTAGTAGT     566

    25/25  X. laevis    542    TAGTAGTAGTAGTAGTAGTAGTAGT     566
    1.00   Sea urchin   107    TAGTAGTAGTAGTAGTAGTAGTAGT     83


                                                T             G
                                                ↑             ↑
c)  28/35  LaBelle      3863   TGTGGCGTCTCACAATTGTGTTGTTGTTGTTATTG     3897
                                  *  **  * *    *             *
    0.80   Rat          12     TGTTGCTACTGAGAATTTTGTTGTTGTTGTTGTTG     46

    25/35  Rat          12     TGTTGCTACTGAGAATTTTGTTGTTGTTGTTGTTG     46
                                 ** ***  * * ** *
    0.71   Eco6         1274   TGAGGTATCTCACAGGTGTGTTGTTGTTGTTGTTG     1308
                                                ↓
                                                T


d)  LaBelle         3574   5'  TA-AATATACTT   3'   3564
                                 *     *
    T. glabrata            5'  TATAATATTCTT   3'   Consensus
                                *    * *
    Eco6            838    5'  TCTAATCTACTT   3'   827
```

Figure 9.

Regions of identity between the putative 1,151 amino acid polypeptide of the LaBelle mitochondrial plasmid and various RNA polymerases.

Numbers in the narrow part of the hat shaped boxes indicate the size of amino acid block under consideration.
Lower case letters in the wide part of the hat shaped boxes refer to the regions shown in Table 11 and the numbers indicate the percentage identity.
The regions that are colinear are joined by diagonal lines.


LaBelle 1151 amino acid
100 amino acids
57
e=35.9
34
j=31.4
49
a=23.2
59
d=36.1
42
f=28.9
49
h=21.8
43
i=16.7
48
n=22.0
45
p=30.0
j=31.4
35
h=21.8
53
i=16.7
48
a=23.2
51
d=36.1
59
f=28.9
42
e=35.9
63
n=22.0
48
p=30.0
47
β' subunit 1407 amino acid
LaBelle 1151 amino acid
43
c=27.9
46
b=34.8
47
g=25.0
47
k=20.0
44
m=26.1
43
o=28.9
48
q=25.0
58
t=23.0
47
s=21.6
b=34.8
46
g=25.0
47
m=26.1
44
o=28.9
44
c=27.9
43
s=21.6
49
k=20.0
50
t=23.0
61
q=25.0
48
mt pol 1351 amino acid

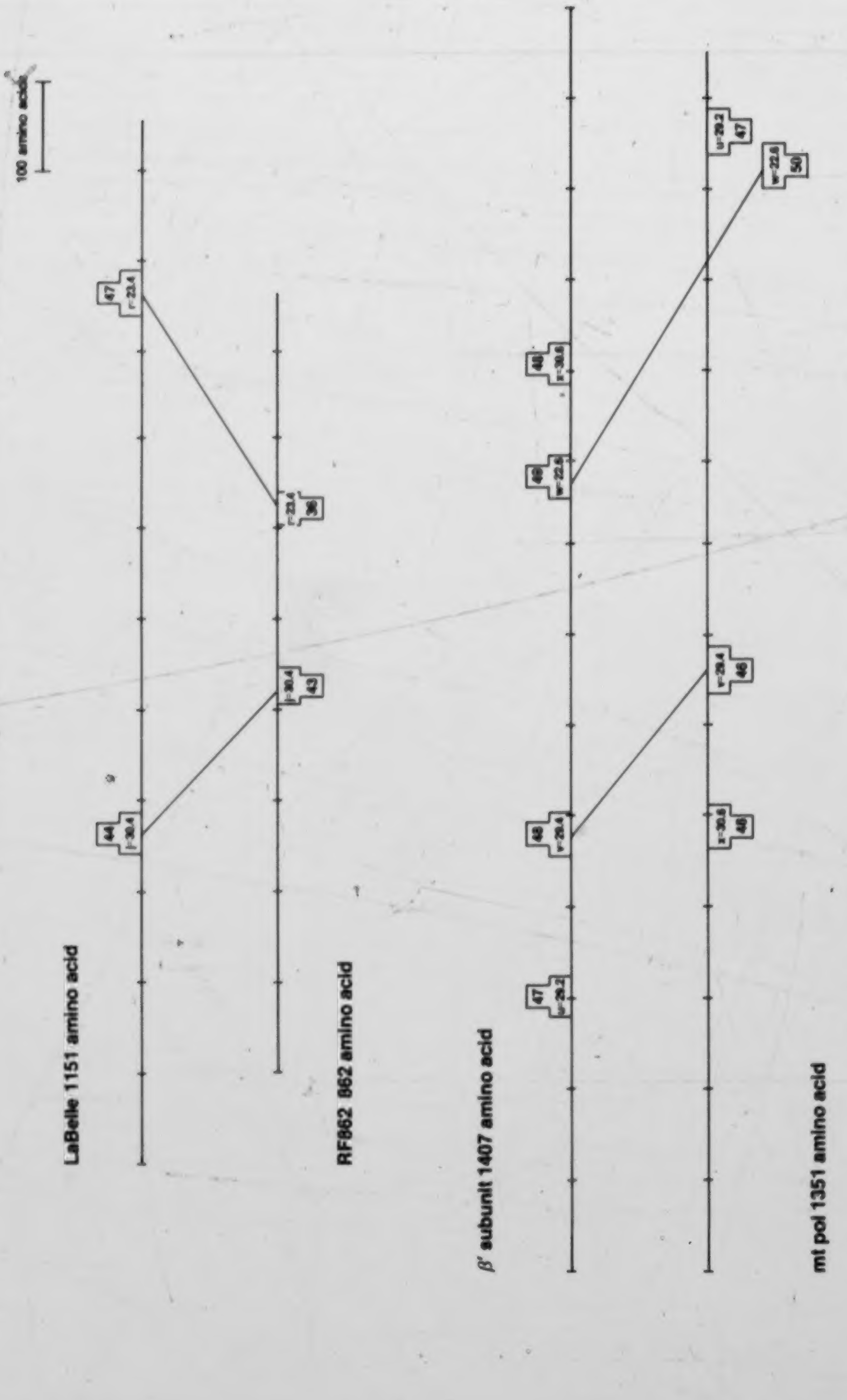

100 amino acid
LaBelle 1151 amino acid
44
j=30.4
47
r=23.4
RF862 862 amino acid
j=30.4
43
r=23.4
36
β' subunit 1407 amino acid
47
u=29.2
48
v=29.4
48
w=22.6
48
x=30.6
mt pol 1351 amino acid
x=30.6
48
v=29.4
46
w=22.6
50
u=29.2
47

Figure 10.

Sequence of the DNA binding domain of λ repressor, λ cro and the putative protein encoded by the LaBelle long ORF.

Data compiled from Pabo and Sauer, 1984.
Residues underlined twice are predicted to make contacts with the DNA backbone.
Residues underlined four times are predicted to make major groove contacts.
Vertical lines indicate positions of surface mutations that decrease operator binding
H indicates hydrophobic amino acid.
Numbers at the sides indicate amino acid residue position in the respective protein.
Crosses indicate the amino acid residues in the DNA binding domain of the LaBelle long ORF protein which are different from the consensus sequences.
Helix 2 and 3 refer to the α-helix of the protein.

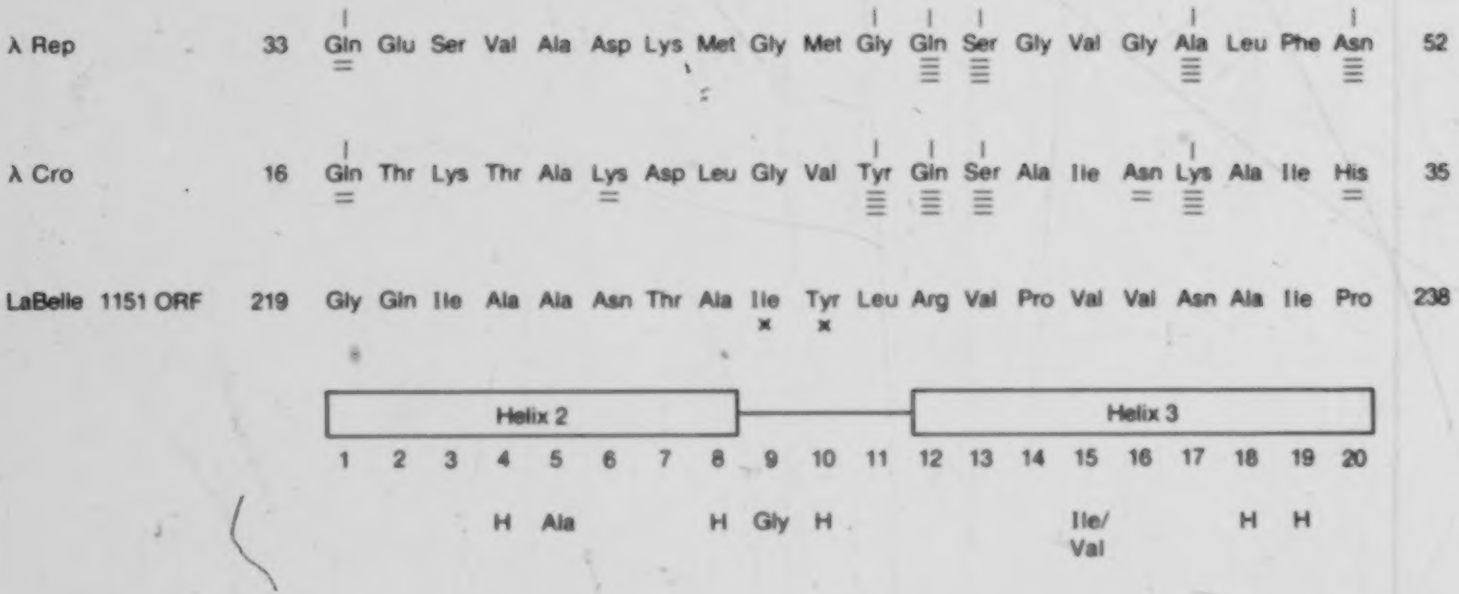
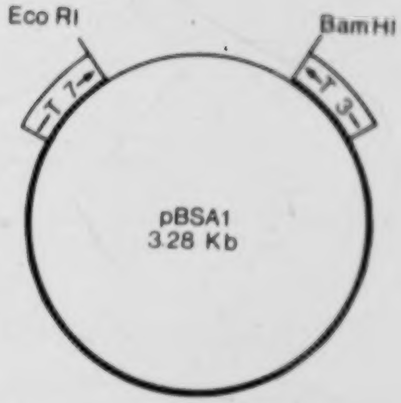

λ Rep 33 Gln Glu Ser Val Ala Asp Lys Met Gly Met Gly Gln Ser Gly Val Gly Ala Leu Phe Asn 52
λ Cro 16 Gln Thr Lys Thr Ala Lys Asp Leu Gly Val Tyr Gln Ser Ala Ile Asn Lys Ala Ile His 35
LaBelle 1151 ORF 219 Gly Gln Ile Ala Ala Asn Thr Ala Ile Tyr Leu Arg Val Pro Val Val Asn Ala Ile Pro 238
x
x
Helix 2
Helix 3
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
H
Ala
H
Gly
H
Ile/
Val
H
H

Figure 11.

Recombinant plasmids pBSA1 and pBSA15 constructed to prepare single stranded riboprobes and cRNA.

The thick lines represents the 2.725 Kb Bluescribe vector.
The thin lines represents the 555 bp insert in pBSA1 or the 3.515 Kb insert in pBSA15.
T3: Orientation of T3 RNA polymerase promoter.
T7: Orientation of T7 RNA polymerase promoter.
The arc inside the pBSA15 plasmid indicates the region containing the 1,151 amino acid ORF with its amino and carboxy termini labelled.
The NH_2 terminus corresponds to nucleotides 59 to 61 of the LaBelle plasmid. The COOH terminus corresponds to nucleotides 3512 to 3514 of the LaBelle plasmid.
The EcoRI site defines nucleotide 1 of the LaBelle plasmid.
The BamHI site begins at nucleotide 3516 of the LaBelle plasmid.


Eco RI
BamHI
T 7
T 3
pBSA1
3.28 Kb

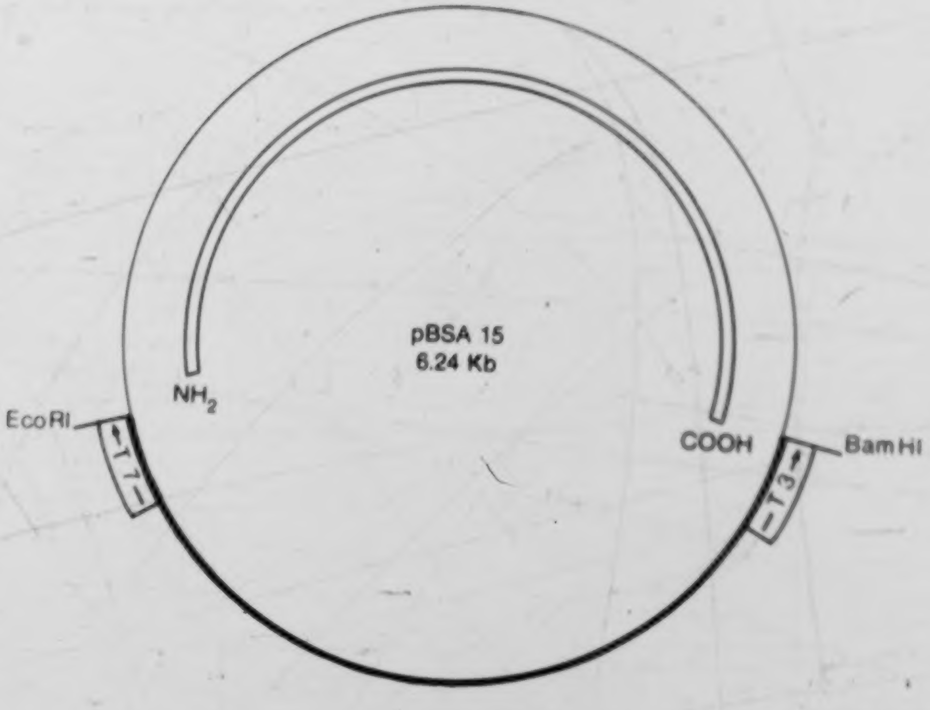

pBSA 15
6.24 Kb
NH$_2$
COOH
EcoRI
T 7
T 3
BamHI

Figure 12.

Map of *N. crassa* mtDNA.

The gene organization depicted here is assembled from data published by Burger and Werner, 1986; Nelson and Macino, 1987; de Vries *et al.*, 1986 and references therein.

Abbreviations:

CO1, CO2, and CO3 genes for cytochrome *c* oxidase subunits 1, 2, and 3 respectively.
URFA1 a previous unassigned open reading frame with homology to the ND6 gene of animals (Burger and Werner, 1986).
COB, gene for apocytochrome b.
URFU and URFN, unidentified reading frames flanking the CO1 gene.
ND1, ND2, ND3, ND4L, ND5 and ND6, genes homologous to corresponding genes in mammalian mitochondria that encode components of NADH dehydrogenase.
ATPase 6 and 8, subunits 6 and 8 of the mitochondrial ATP synthase.
MAL, an open reading frame with homology to subunit 9 of the ATP synthase which is encoded by a nuclear gene in *N. crassa* and translated in the cytoplasm.
s-rRNA, L-rRNA, small and large ribosomal RNA.
tRNAs are marked by thin bars.
Exons are indicated by filled blocks.
Unassigned and intronic reading frames are shaded.
ND3 is shown by broken lines as only part of the sequence was available at the time of preparation of this figure (de Vries *et al.*, 1986).
The inner ring represents the EcoRI restriction map.
The position of KpnI restriction sites are indicated on the EcoRI-6 mtDNA fragment. The homology between the LaBelle mitochondrial plasmid and the EcoRI-6 mtDNA fragment includes the KpnI internal fragment and extends to near the junction of the EcoRI-4 and EcoRI-6 mtDNA fragments.

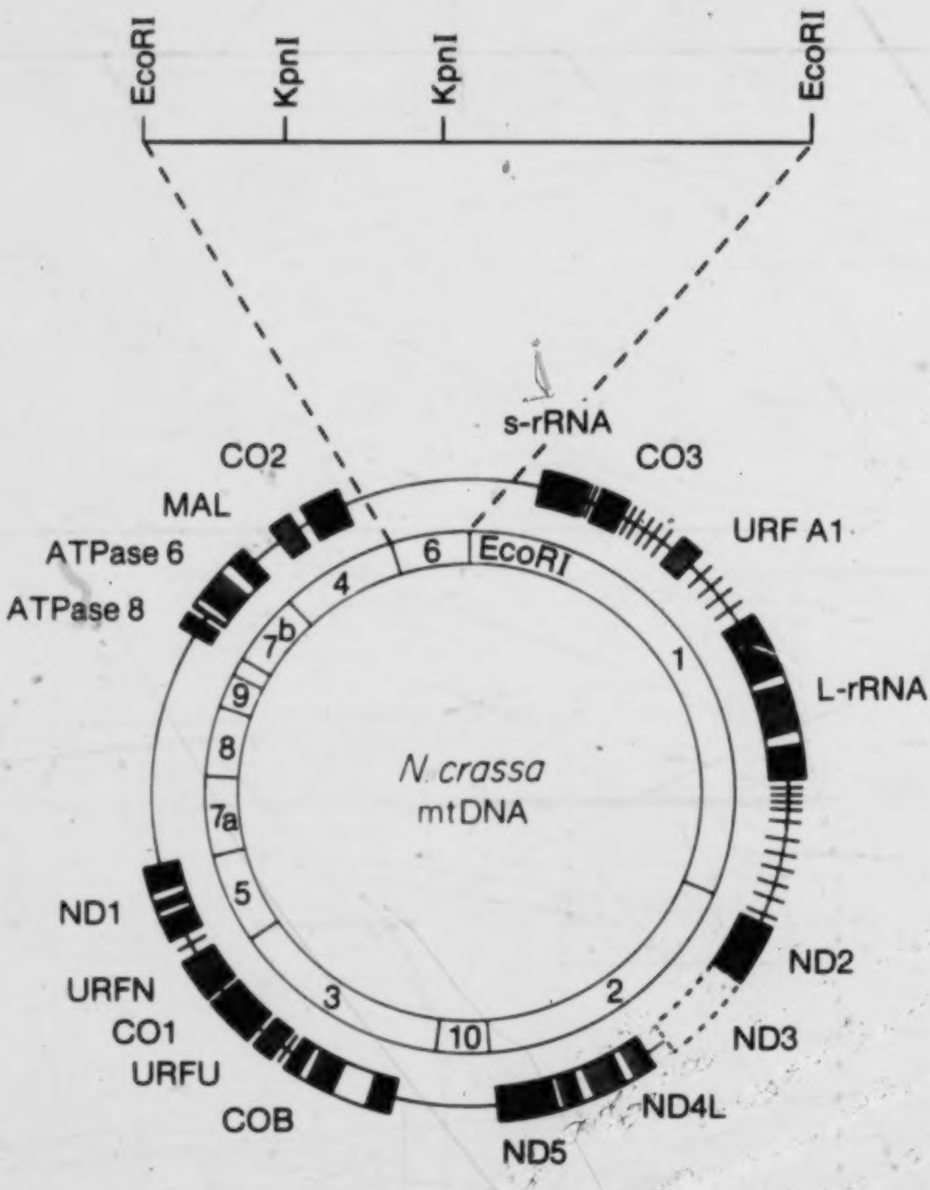

EcoRI
KpnI
KpnI
EcoRI
s-rRNA
CO3
URF A1
L-rRNA
ND2
ND3
ND4L
ND5
COB
URFU
CO1
URFN
ND1
ATPase 8
ATPase 6
MAL
CO2
6
EcoRI
4
7b
9
8
7a
5
3
10
2
1
N. crassa
mtDNA

Figure 13.

Linear Map of the EcoRI-6 fragment of *Neurospora* mtDNA showing the region homologous to the LaBelle mitochondrial plasmid DNA.

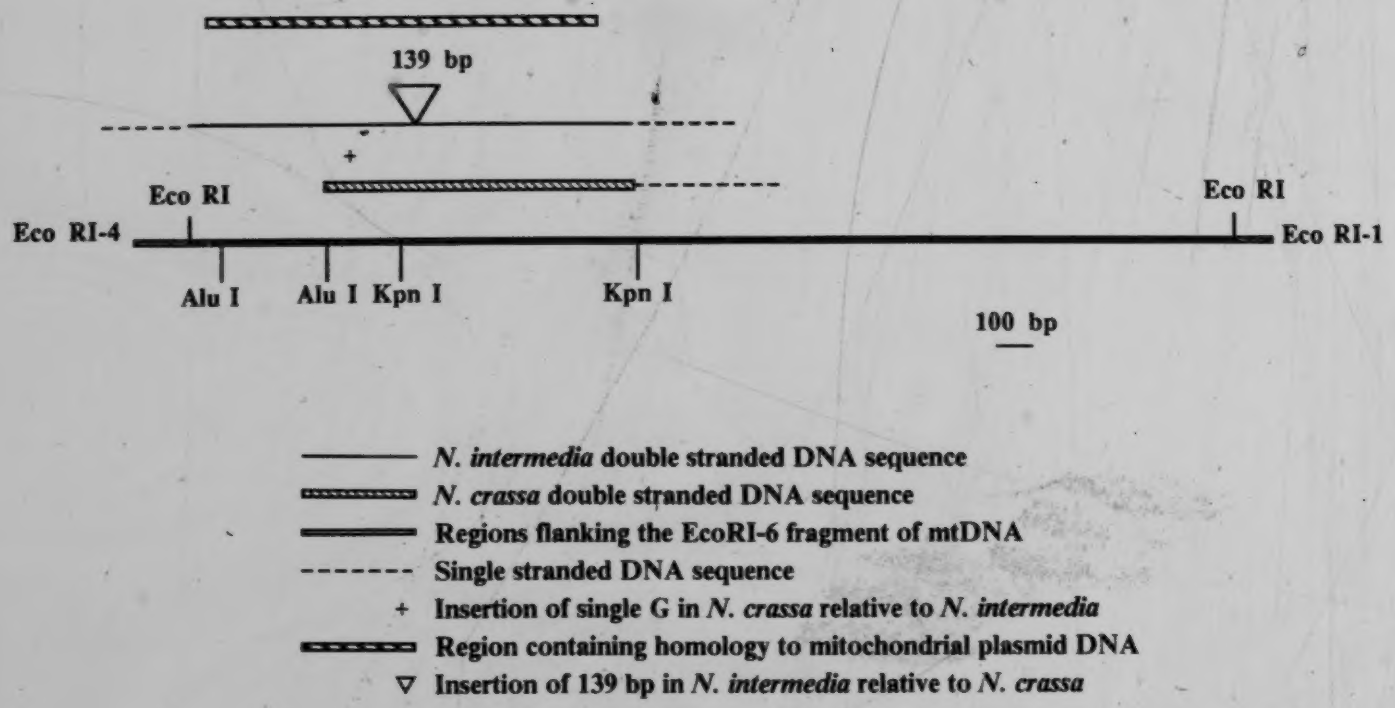

139 bp
+
Eco RI
Eco RI-4
Alu I
Alu I
Kpn I
Kpn I
Eco RI
Eco RI-1
100 bp
N. intermedia double stranded DNA sequence
N. crassa double stranded DNA sequence
Regions flanking the EcoRI-6 fragment of mtDNA
Single stranded DNA sequence
+ Insertion of single G in N. crassa relative to N. intermedia
Region containing homology to mitochondrial plasmid DNA
▽ Insertion of 139 bp in N. intermedia relative to N. crassa

Figure 14.

Sequencing strategy used for the EcoRI-6 fragment of *N. intermedia* strain LaBelle-1b mtDNA.

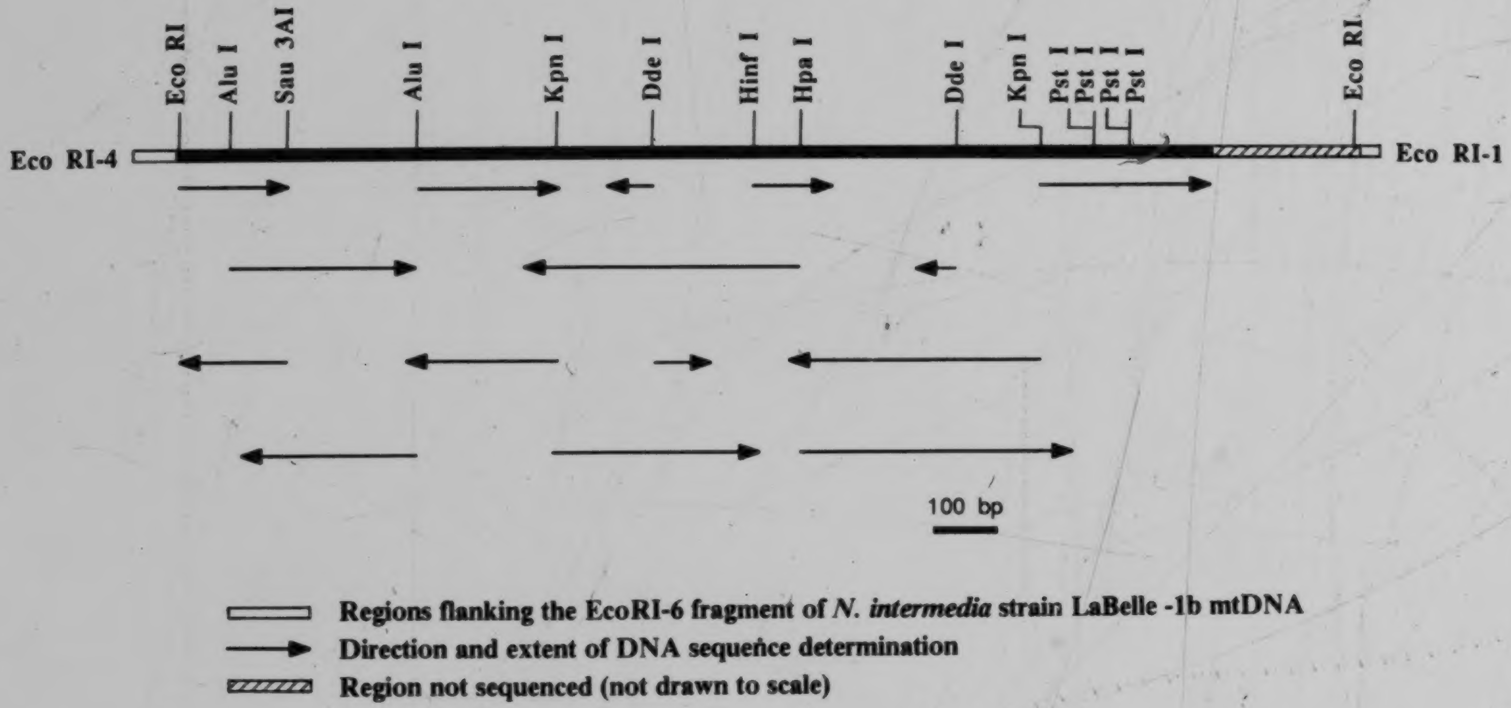

Eco RI
Alu I
Sau 3AI
Alu I
Kpn I
Dde I
Hinf I
Hpa I
Dde I
Kpn I
Pst I
Pst I
Pst I
Pst I
Eco RI
Eco RI-4
Eco RI-1
100 bp
Regions flanking the EcoRI-6 fragment of *N. intermedia* strain LaBelle -1b mtDNA
Direction and extent of DNA sequence determination
Region not sequenced (not drawn to scale)

Figure 15.

DNA sequence of a portion of the EcoRI-6 fragment of *N. intermedia* strain LaBelle-1b mtDNA.

Both strands of the DNA were sequenced from nucleotide 1 to nucleotide 1488. The top strand alone was sequenced from nucleotide 1489 to nucleotide 1774. The sequence of the bottom strand was deduced from this information.
The first nucleotide in the EcoRI site at the EcoRI-4/6 junction has been arbitrarily assigned position 1.
The numbers on top of every 10th residue refer to nucleotide positions in the EcoRI-6 mtDNA fragment.
Nucleotides 1559 to 1572 inclusive form the first perfect PstI palindromic sequence and nucleotides 1632 to 1645 inclusive form the second perfect PstI palindromic sequence.

```
        10         20         30         40         50         60
GAATTCTTAG TTTATGAGTT TAGGAAAAGG ATTTAAAGTA TTGATGGGAT TAAAAGATTT
CTTAAGAATC AAATACTCAA ATCCTTTTCC TAAATTTCAT AACTACCCTA ATTTTCTAAA

        70         80         90        100        110        120
GAGTATATTG AAAATGGTTT ACGTAAAGCT ATAGATATTT ATGAAAAAGG GGGTCATGTG
CTCATATAAC TTTTACCAAA TGCATTTCGA TATCTATAAA TACTTTTTCC CCCAGTACAC

       130        140        150        160        170        180
CTACCTAGTT ATCGTGATAA GAGTTTTAGT ATAGAGTACG ACAATAAGCG TCGTGTTGTT
GATGGATCAA TAGCACTATT CTCAAAATCA TATCTCATGC TGTTATTCGC AGCACAACAA

       190        200        210        220        230        240
AGTCAATGAT CTATAAGTAG TACTATGTAT GATACAGTGC CTTGGTGTAC GATAAATGAT
TCAGTTACTA GATATTCATC ATGATACATA CTATGTCACG GAACCACATG CTATTTACTA

       250        260        270        280        290        300
TATGGGCCCC CCGCCCCGAA GGGGGCGGGG TCCCCCGAAA AATAAGGATA CTAAAAGAAA
ATACCCGGGG GGCGGGGCTT CCCCCGCCCC AGGGGGCTTT TTATTCCTAT GATTTTCTTT

       310        320        330        340        350        360
CGGTAAATAC GCCTGTATTT GTAAAAGGAG CGTACGCTGT AGGGGGTGGA AAAGGTTATA
GCCATTTATG CGGACATAAA CATTTTCCTC GCATGCGACA TCCCCCACCT TTTCCAATAT

       370        380        390        400        410        420
AAAGTACTAT AGAAACCACC CCTCCGGGGT GGATGCATTC ATTAAAGCAG CTTTCTCCGA
TTTCATGATA TCTTTGGTGG GGAGGCCCCA CCTACGTAAG TAATTTCGTC GAAAGAGGCT

       430        440        450        460        470        480
TAGTGAACGT TTTGGTATAC CTGAAAATGA GTTTAAAACT TATAATTCCA CTCCACTCCC
ATCACTTGCA AAACCATATG GACTTTTACT CAAATTTTGA ATATTAAGGT GAGGTGAGGG

       490        500        510        520        530        540
CCCCCCCTTA AGGGGGGGGA GTGGATTTTA GACGCCTTGC GGCTCTGACC AAGCAAAGAA
GGGGGGGAAT TCCCCCCCCT CACCTAAAAT CTGCGGAACG CCGAGACTGG TTCGTTTCTT

       550        560        570        580        590        600
AGTAAAACTT ACAAAAACTA GTATTTCGGG GGACCCCGCT TCGGGGGCGG GGGCCCAAGT
TCATTTTGAA TGTTTTTGAT CATAAAGCCC CCTGGGGCGA AGCCCCCGCC CCCGGGTTCA

       610        620        630        640        650        660
TAAAATGTCG TAAAACGATA TCGAGAACGG TACCTCGGAC AAAAGAAACA GAAGGATTTA
ATTTTACAGC ATTTTGCTAT AGCTCTTGCC ATGGAGCCTG TTTTCTTTGT CTTCCTAAAT

       670        680        690        700        710        720
TAGAGTATGT AAAGAAACGA ATAAGTACAT TCGATAGTGA TCGTTTCTTT AAGAAACGTA
ATCTCATACA TTTCTTTGCT TATTCATGTA AGCTATCACT AGCAAAGAAA TTCTTTGCAT

       730        740        750        760        770        780
ATAACGATGA AGATTAAAAT GCCTCTTTAA GTATAAGTTT TATCTGATAG AAATATTTAA
TATTGCTACT TCTAATTTTA CGGAGAAATT CATATTCAAA ATAGACTATC TTTATAAATT

       790        800        810        820        830        840
CGCCCTTATC AATAATATAC TGAGTAACAG TTCAATTCGT ATTTTTAAGT AGATTAGAGC
GCGGGAATAG TTATTATATG ACTCATTGTC AAGTTAAGCA TAAAAATTCA TCTAATCTCG

       850        860        870        880        890        900
CGCAAGGCGT CTGTGAATTA AATGTAAATG TAAGCTATTT TTTGATAAAA TTGTTAGAGG
GCGTTCCGCA GACACTTAAT TTACATTTAC ATTCGATAAA AAACTATTTT AACAATCTCC
```

```
       910        920        930        940        950        960
TGTATAAGCG CTCGTTTAGT TTGCTTTGCT TTGTGTGTAA AAAAGTCCAC ATTTTTTAGA
ACATATTCGC GAGCAAATCA AACGAAACGA AACACACATT TTTTCAGGTG TAAAAAATCT

       970        980        990       1000       1010       1020
TACGAAAAAA TGTGGTAAAA AACCGGATTC CACCCCGGAG GGGTGGACGG TAACTGTGAC
ATGCTTTTTT ACACCATTTT TTGGCCTAAG GTGGGGCCTC CCCACCTGCC ATTGACACTG

      1030       1040       1050       1060       1070       1080
AATTATAGAC ATGCGGGTGT GTGGGACCAT GACATTACGT CAGGTTAACG AAGTATCAGT
TTAATATCTG TACGCCCACA CACCCTGGTA CTGTAATGCA GTCCAATTGC TTCATAGTCA

      1090       1100       1110       1120       1130       1140
AAAAATAATA ATAGTATGTT TAAATATTAT AATGTATGTG TGGTAGTAGT AATAATAATA
TTTTTATTAT TATCATACAA ATTTATAATA TTACATACAC ACCATCATCA TTATTATTAT

      1150       1160       1170       1180       1190       1200
TTATTAAATA TTTAATTATT TTTAACTTCG GGAATTTTTA ATTATCAAAT AAGAATATGA
AATAATTTAT AAATTAATAA AAATTGAAGC CCTTAAAAAT TAATAGTTTA TTCTTATACT

      1210       1220       1230       1240       1250       1260
AGGATGCAAC GGTTTATTAG TGTAAGAAAT ATTTTATTAT TTTAAAATTA AAAAACAGTA
TCCTACGTTG CCAAATAATC ACATTCTTTA TAAAATAATA AAATTTTAAT TTTTTGTCAT

      1270       1280       1290       1300       1310       1320
TGTTTACAAC CTGTGAGGTA TCTCACAGGT GTGTTGTTGT TGTTGTTGTA GTTGTTGTTG
ACAAATGTTG GACACTCCAT AGAGTGTCCA CACAACAACA ACAACAACAT CAACAACAAC

      1330       1340       1350       1360       1370       1380
CTTATAAGTA CTAAGTGCTA ACTGCTGCCT ACTTAAATAG TTTGGTAAAG TATTTATCCT
GAATATTCAT GATTCACGAT TGACGACGGA TGAATTTATC AAACCATTTC ATAAATAGGA

      1390       1400       1410       1420       1430       1440
CACATCGTAA GGTGTGGGTA GGATTAAAGT TTGAGTATTG GGCGGATATC CTCGCAACGT
GTGTAGCATT CCACACCCAT CCTAATTTCA AACTCATAAC CCGCCTATAG GAGCGTTGCA

      1450       1460       1470       1480       1490       1500
TATTTATAAT CCATGTGACG AATACCCAAC TCTCAAGGCA TGGGTACCTG ATAGAGTGGG
ATAAATATTA GGTACACTGC TTATGGGTTG AGAGTTCCGT ACCCATGGAC TATCTCACCC

      1510       1520       1530       1540       1550       1560
CCTATTTGTG CTTAGTAGGC ATCGGCGTGA GTCTAAATCG CAACCCCCTC CTTTCCCCCT
GGATAAACAC GAATCATCCG TAGCCGCACT CAGATTTAGC GTTGGGGGAG GAAAGGGGGA

      1570       1580       1590       1600       1610       1620
GCAGTACTGC AGGGGGGTGG GGAAGGGGAG GGGGAAGAGG AGGGGGGATT TGCCTTTTTC
CGTCATGACG TCCCCCCACC CCTTCCCCTC CCCCTTCTCC TCCCCCCTAA ACGGAAAAAG

      1630       1640       1650       1660       1670       1680
GTTCCCCCGC CCTGCAGTAC TGCAGGGGGG GGTGTAATAG AAATATTATA TAAGTGCTAA
CAAGGGGGCG GGACGTCATG ACGTCCCCCC CCACATTATC TTTATAATAT ATTCACGATT

      1690       1700       1710       1720       1730       1740
GAGGTAGGAC GGTTCGAATC CGGAGGCTAA GAATATTTAA ATTTAAATCC CCCCCTCCTC
CTCCATCCTG CCAAGCTTAG GCCTCCGATT CTTATAAATT TAAATTTAGG GGGGGAGGAG

      1750       1760       1770
TTACGTATCC GAAGGGAGGC ACGCCCCCTC AAGG
AATGCATAGG CTTCCCTCCG TGCGGGGGAG TTCC
```

Figure 16.

Nucleotide sequence of homologous regions between the LaBelle mitochondrial plasmid and the *N. intermedia* EcoRI-6 mtDNA fragment.

Numbers correspond to the numbering in Figs. 2 and 15.

A C G T : Matched nucleotides

A C G T : Mismatches

A C G T : Insertion

_ _ _ _ : Deletion

▾ ▴ : Position of nucleotide numbered

```
              2900                2910                2920                2930
               ▼                   ▼                   ▼                   ▼
                       A             A       G         G T C   C
LaBelle    T T A A A A   A T T T G A   T A T   T T G A       T   G T T T A C G T A A A G C
N.i.Eco 6  T T A A A A   A T T T G A   T A T   T T G A       T   G T T T A C G T A A A G C
                       G             G       A         A A A   G
           ▲                   ▲                   ▲                   ▲
           50                  60                  70                  80

              2940                2950                2960                2970
               ▼                   ▼                   ▼                   ▼
               A G                 C     T                             T     G A T   G A G
LaBelle    T A     G A T A T T T A   G A   A A A G G G G G T C A T G T   C T       T       T A
N.i.Eco 6  T A     G A T A T T T A   G A   A A A G G G G G T C A T G T   C T       T       T A
               T A                 T     A                             G     A C C   A G T
           ▲                   ▲                   ▲                   ▲                   ▲
           90                  100                 110                 120                 130

          2980                2990                3000                3010                3020
           ▼                   ▼                   ▼                   ▼                   ▼
           C A   C     C                             T           T     T           A
LaBelle        G   G A   A A G A G T T T T A G T A T   G A G T A   G A   A A T A A   C G T C
N.i.Eco 6      G   G A   A A G A G T T T T A G T A T   G A G T A   G A   A A T A A   C G T C
           T C   T     T                             A           C     C           G
                           ▲                   ▲                   ▲                   ▲
                          140                 150                 160                 170
```

```
                            3030                3040                3050                3060
                             ▼                   ▼                   ▼                   ▼
               A         C   A   G C T           G G   G T     G           C       T A     G T
LaBelle    G T   T T G T   A   T       T G A T C     T     G T   G T A C T   T G T     G A
N.i.Eco 6  G T   T T G T   A   T       T G A T C     T     G T   G T A C T   T G T     G A
               G         T   G   C A A           T A   A A     A           A       A T     T A
                         ▲                   ▲                   ▲                   ▲
                        180                 190                 200                 210

                        3070                3080                3090
                         ▼                   ▼                   ▼
             T     A           A G     T             C   C C         A - - - - - - - - - - - - -
LaBelle    C   G T   C C T T G     G T   C G A T A A   T     T T A T
N.i.Eco 6  C   G T   C C T T G     G T   C G A T A A   T     T T A T
             A     G           G T     A             A   G A         G G G C C C C C C G C C C C
                     ▲                   ▲                   ▲                   ▲
                    220                 230                 240                 250

                                                                  3100                3110
                                                                   ▼                   ▼
           - - - - - - - - - - - - - - - - - - -           G                 A T T         G
LaBelle                                              G A A   A A T A A G G A       T A A A   G A A A
N.i.Eco 6                                            G A A   A A T A A G G A       T A A A   G A A A
           G A A G G G G G C G G G G T C C C C C           A                 T A C         A
               ▲                   ▲                       ▲                   ▲                   ▲
              260                 270                     280                 290                 300
```

```
               3120                3130                3140                3150
                ▼                   ▼                   ▼                   ▼
            C     T           T     A           C                       C   T T A       G T
LaBelle   C   G T   A A T A C   C C   G T A T T   G T A A A A G G A G C   T       C T G
N.i.Eco 6 C   G T   A A T A C   C C   G T A T T   G T A A A A G G A G C   T       C T G
            G     A           G     T           T                       G   A C G       T A
                            ▲                   ▲                   ▲                   ▲
                            310                 320                 330                 340

             3160                3170                3180                3190
              ▼                   ▼                   ▼                   ▼
                A                 A A G                         T           T A G T G T
LaBelle   G G G   G T G G A A A A       T A T A A A A G T A C T A T   G A A A C
N.i.Eco 6 G G G   G T G G A A A A       T A T A A A A G T A C T A T   G A A A C
                G                 G G T                         A           C A C C C C
                          ▲                   ▲                   ▲                   ▲
                          350                 360                 370                 380

           3200                                3210                3220                3230
            ▼                                   ▼                   ▼                   ▼
          T A G A G G T - - - - - - - -                       T       C   A     T         A
LaBelle                                   T T C A T T A A A G C   G C T   T   T C   G A T A   T
N.i.Eco 6                                 T T C A T T A A A G C   G C T   T   T C   G A T A   T
          T C C G G G G T G G A T G C A                       A       T   C     C         G
                        ▲                   ▲                   ▲                   ▲
                        390                 400                 410                 420
```

```
                       3240                3250                3260                3270
                        ▼                   ▼                   ▼                   ▼
              C                       T     C A     G     C   A       C
LaBelle   G A   C G T T T T G G T A T   C C     A A   A T   A   T T T   A A A C T T A T A A
N.i.Eco 6 G A   C G T T T T G G T A T   C C     A A   A T   A   T T T   A A A C T T A T A A
              A                       A     T G     A     G   G       A
                    ▲                   ▲                   ▲                   ▲
                   430                 440                 450                 460

                     3280                3290                3300
                      ▼                   ▼                   ▼
            G A T A T C A T T A A A T T C A T A C G C G G T C A - - - - - - - - - - - - -
LaBelle   T
N.i.Eco 6 T
            T C C A C T C C A C T C C C C C C C C C C T T A A G G G G G G G G A G T G G A
                  ▲                   ▲                   ▲                   ▲
                 470                 480                 490                 500

                                                                   3310                3320
                                                                    ▼                   ▼
          - - - - - - - - - - - - - - - - - - - -       T G C       C           T     T
LaBelle                                           T G A       A G C   A A G A A   G T   A A A C T
N.i.Eco 6                                         T G A       A G C   A A G A A   G T   A A A C T
          T T T T A G A C G C C T T G C G G C T C       C C A       A           A     A
                  ▲                   ▲                   ▲                   ▲
                 510                 520                 530                 540
```

```
                   3330                3340
                    ▼                   ▼
          C T         G                     - - - - - - - - - - - - - - - - - - - - - - - - -
LaBelle       C A A A   A C T A G T A T T T
N.i.Eco 6     C A A A   A C T A G T A T T T
          T A         A                     C G G G G G A C C C C G C T T C G G G G G C G G G
          ▲                   ▲                   ▲                   ▲                   ▲
         550                 560                 570                 580                 590

                               3350                3360                3370                3380
                                ▼                   ▼                   ▼                   ▼
          - - -         T                             T G   T       G   A C
LaBelle         C C A A   T T A A A A T G T C G T A A     C   A T A   C     G A A C G G T A C C T
N.i.Eco 6       C C A A   T T A A A A T G T C G T A A     C   A T A   C     G A A C G G T A C C T
          G G C         G                             A A   G       T   G A
                          ▲                   ▲                   ▲                   ▲
                         600                 610                 620                 630

                       3390                3400                3410                3420
                        ▼                   ▼                   ▼                   ▼
              T       G   C                       G             A   A           C   C T
LaBelle   C G   A C A   A   G A A A C A G A A G G   T T T A T A   A   T A T G T   A     A A
N.i.Eco 6 C G   A C A   A   G A A A C A G A A G G   T T T A T A   A   T A T G T   A     A A
              G       A   A                       A             G   G           A   A G
                  ▲                   ▲                   ▲                   ▲
                 640                 650                 660                 670
```

```
                       3430                3440                3450                3460
                        ▼                   ▼                   ▼                   ▼
              A A           C C     C           C G T                                 G
LaBelle     A     A A T A A     A C   T T C G A       T G A T C G T T T C T T T A A G   A A
N.i.Eco 6   A     A A T A A     A C   T T C G A       T G A T C G T T T C T T T A A G   A A
              C G           G T     A           T A G                                 —
                  ▲                   ▲                   ▲                   ▲
                 680                 690                 700                 710

                       3470                3480                3490                3500
                        ▼                   ▼                   ▼                   ▼
              A A C C T C A A G A A A A A G C G A A C A A A A A A G A T A G G A T A G A A A
LaBelle     A
N.I.Eco 6   A
              C G T A A T A A C G A T G A A G A T T A A A A T G C C T C T T T A A G T A T A
                    ▲                   ▲                   ▲                   ▲
                   720                 730                 740                 750

                         3510                3520                3530                3540
                          ▼                   ▼                   ▼                   ▼
              G T G A T A A T T A G A G G A T C C C C C T T T C A A T A T T A T C T T T A A T
LaBelle
N.i.Eco 6
              A G T T T T A T C T G A T A G A A A T A T T T A A C G C C C T T A T C A A T A A
                      ▲                   ▲                   ▲                   ▲
                     760                 770                 780                 790
                    3550                                     3560              3570
                      ▼                                        ▼                 ▼
          A G C A G C T C T  - - - - - - - - - - -     A   -                   T  T T T A
LaBelle                                         A T T C  T  T T T T T A A G T A  A
N.i.Eco 6                                       A T T C  T  T T T T T A A G T A  A
          T A T A C T G A G T A A C A G T T C A        G   A                   G  T T A G
                  ▲                   ▲                  ▲                 ▲
                 800                 810                820               830

                  3580                3590                3600
                    ▼                   ▼                   ▼
          T T A G T A G T G T A T C G T T A G A A A A A A A T C - - - - - - - - - - - - -
LaBelle
N.i.Eco 6
          A G C C G C A A G G C G T C T G T G A A T T A A A T G T A A A T G T A A G C T A
              ▲                   ▲                   ▲                   ▲
             840                 850                 860                 870

          - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
LaBelle
N.i.Eco 6
          T T T T T T G A T A A A A A T T G T T A G A G G T G T A T A A G C G C T C G T T T A
              ▲                   ▲                   ▲                   ▲
             880                 890                 900                 910
```

```
                                                               3610                3620
                                                                ▼                   ▼
          - - - - - - - - - - - - - - - - - - -               G G T                           G
LaBelle                                       T A A A A A A       G T C C A C A T T T T T T
N.i.Eco 6                                     T A A A A A A       G T C C A C A T T T T T T
          G T T T G C T T T G C T T T G T G T G               - - -                           A
            ▲                   ▲                   ▲                         ▲
           920                 930                 940                       950

                 3630                3640                3650                3660
                  ▼                   ▼                   ▼                   ▼
              A     C   T         A     C T A C A T T T T T T G G T T T C T A G A A A A A T
LaBelle   G A   A C   A   A A A A   G T
N.i.Eco 6 G A   A C   A   A A A A   G T
              T     G   A         T     G G T A A A A A A C C G G A T T C C A C C C C G G A
            ▲                   ▲                   ▲                   ▲
           960                 970                 980                 990

                     3670                3680                3690                3700
                      ▼                   ▼                   ▼                   ▼
          - - -                 T A G A T A           A   A     G T     G
LaBelle         G T G G A C G G             T A A C T   T   A C     T T   T A G A C A T G C G G G
N.i.Eco 6       G T G G A C G G             T A A C T   T   A C     T T   T A G A C A T G C G G G
          G G G                 - - - - - -           G   G     A A     A
          ▲                   ▲                               ▲                   ▲
         1000                1010                            1020                1030
```

```
               3710                3720                  3730                3740                  3750
                ▼                   ▼                     ▼                   ▼                     ▼
                G           G                                                             T   T   -
aBelle      T G   G T G G G   A C C A T G A C A T T A C G T C A G G T T A A C G A A G T   T   A   T
.i.Eco 6    T G   G T G G G _ A C C A T G A C A T T A C G T C A G G T T A A C G A A G T   T   A   T
                T           -                                                             A   C   G
                ▲                     ▲                     ▲                   ▲                   ▲
               1040                  1050                  1060                1070                1080

                      A     - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
aBelle      A A A A A   A
.i.Eco 6    A A A A A   A
                      T   A T A A T A G T A T G T T T A A A T A T T A T A A T G T A T G T
                              ▲                   ▲                   ▲
                             1090                1100                1110

            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
aBelle
.i.Eco 6
            G T G G T A G T A G T A A T A A T A A T A T T A T T A A A T A T T T A A T T A
            ▲                   ▲                   ▲                   ▲
           1120                1130                1140                1150
```

```
                    3760                   3770                3780                3790
                        ▼                     ▼                   ▼                   ▼
          - - - - - T                     -         G                 T           G           G
aBelle                 A C T T C G G G A A   T T T T   A T T A T C A A   T A A G A   T A T G A   G
.i.Eco 6               A C T T C G G G A A   T T T T   A T T A T C A A   T A A G A   T A T G A   G
          T T T T T A                     T         A                 A           A           A
            ▲                   ▲                   ▲                   ▲                   ▲
           1160               1170                1180                 1190              1200

                3800                3810               3820                 3830
                 ▼                   ▼                   ▼                   ▼
               T     T G T       A                                       C         - - -   G   T
aBelle   G A T   C A       G T T   A T T A G T G T A A G A A A T A T T T   T A T T       T   A
.i.Eco 6 G A T   C A       G T T   A T T A G T G T A A G A A A T A T T T   T A T T       T   A
               G     A C G       T                                       -         A T T   T   A
                       ▲                   ▲                   ▲                     ▲
                     1210                1220                1230                  1240

              3840                    3850                3860                3870
               ▼                       ▼                   ▼                   ▼
                   G   T     - -   C                         A       T     C G               A T
aBelle   A A T T A   A   A A     G   A T G T T T A C A A C C   G T G   G G     T C T C A C A
.i.Eco 6 A A T T A   A   A A     G   A T G T T T A C A A C C   G T G   G G     T C T C A C A
                   A   A     C A   T                         T       A     T A               G G
                 ▲                   ▲                   ▲                   ▲
                1250                1260                1270                 1280
           3880                3890                3990                3910
            ▼                   ▼                   ▼                   ▼
                                        A           A   C G C A T T A G T A G T A G T A G T
Belle     T G T G T T G T T G T T G T T   T T G T A   T
i.Eco 6   T G T G T T G T T G T T G T T   T T G T A   T
                                        G           G   - - - - - - - - - - - - - - - - - -
          ▲                   ▲                   ▲
         1290                1300                1310

         3920                  3930                3940                3950                3960
          ▼                     ▼                   ▼                   ▼                   ▼
          A G C G G T A G     A       T       T   T         G     C     C C     C       G
Belle                     T   G T T   G T T   G   T T A T   A G   A C     A G   G C T   A C T
i.Eco 6                   T   G T T   G T T   G   T T A T   A G   A C     A G   G C T   A C T
          - - - - - - - -     -       -       -   C         A     T     T A     T       A
                                                ▲                   ▲                   ▲
                                               1320                1330                1340

                         3970                3980
                          ▼                   ▼
                A       G G G C T
Belle     G C T   C C T           A C T T A A A T A
i.Eco 6   G C T   C C T           A C T T A A A T A
                G       - - - - -
                      ▲                           ▲
                     1350                        1359
```

Figure 17.

Analysis of the region of homology between the LaBelle mitochondrial plasmid and the *N. intermedia* EcoRI-6 fragment.

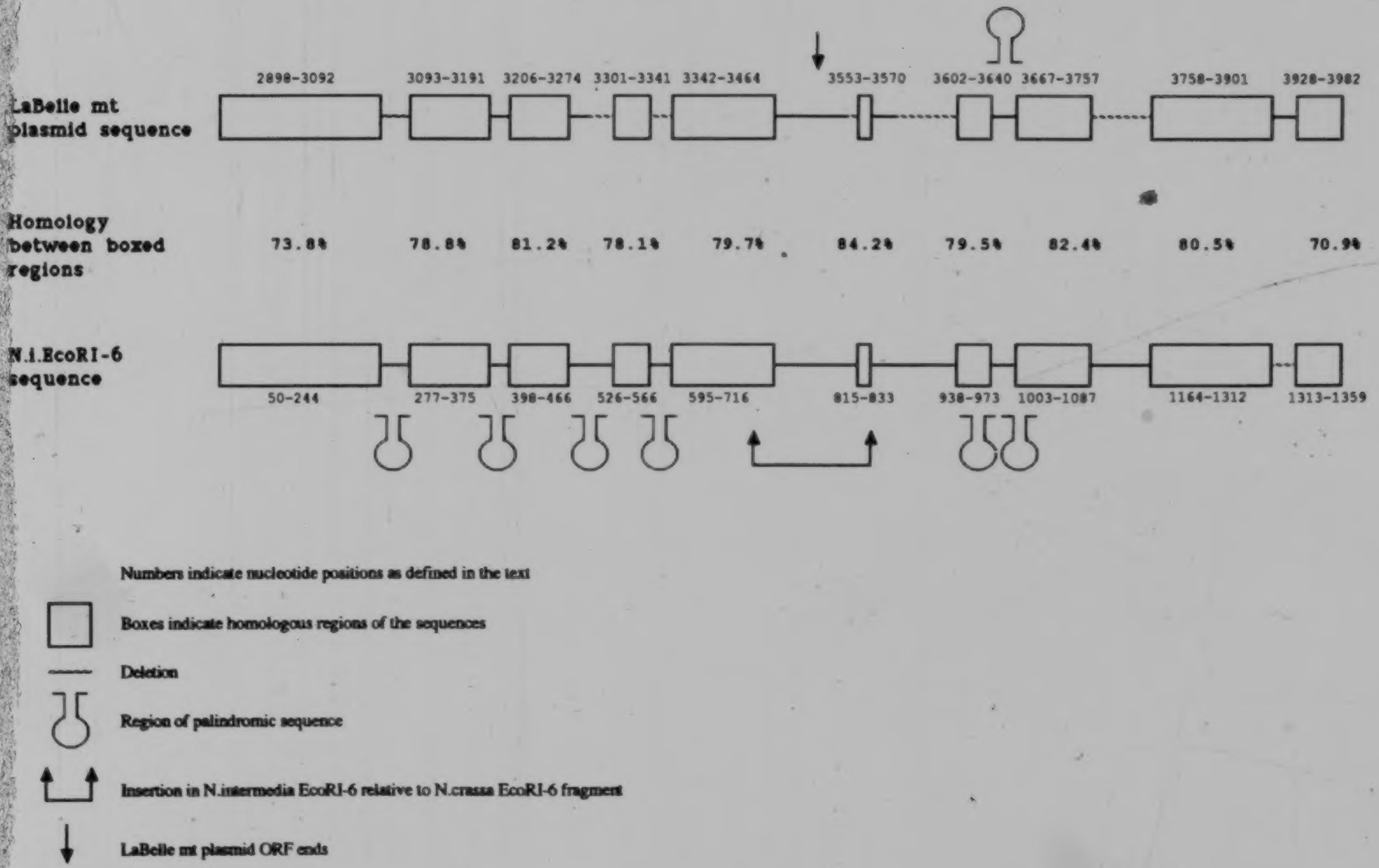

LaBelle mt plasmid sequence
2898-3092
3093-3191
3206-3274
3301-3341
3342-3464
3553-3570
3602-3640
3667-3757
3758-3901
3928-3982
Homology between boxed regions
73.8%
78.8%
81.2%
78.1%
79.7%
84.2%
79.5%
82.4%
80.5%
70.9%
N.i.EcoRI-6 sequence
50-244
277-375
398-466
526-566
595-716
815-833
938-973
1003-1087
1164-1312
1313-1359
Numbers indicate nucleotide positions as defined in the text
Boxes indicate homologous regions of the sequences
Deletion
Region of palindromic sequence
Insertion in N.intermedia EcoRI-6 relative to N.crassa EcoRI-6 fragment
LaBelle mt plasmid ORF ends

Figure 18.

Palindromic sequences capable of forming hairpin loop structures with free energies of formation (ΔG) equal to or less than -10 Kcal/mol in the EcoRI-6 fragment of *N. intermedia* mtDNA .

Numbers shown correspond to numbering in Fig. 15. Unpaired regions are shown as loops.

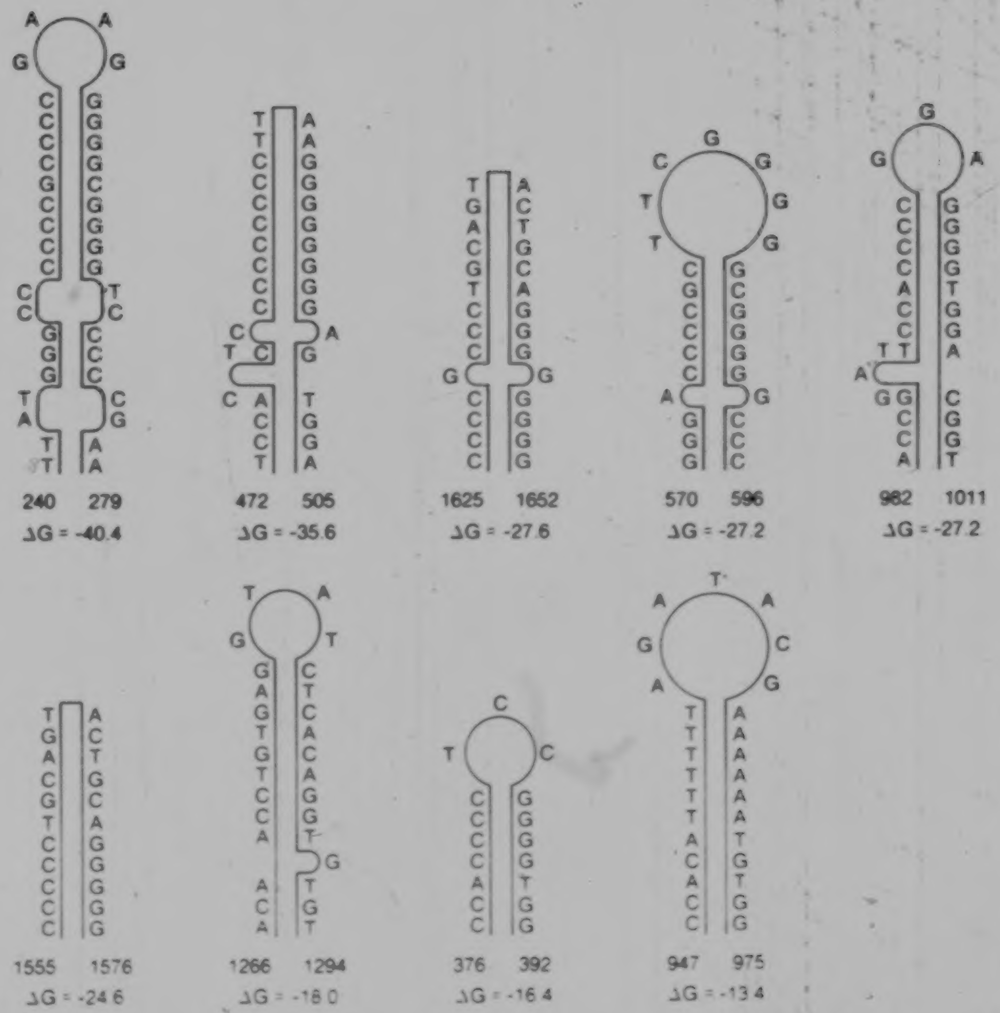

240 279
ΔG = -40.4
472 505
ΔG = -35.6
1625 1652
ΔG = -27.6
570 596
ΔG = -27.2
982 1011
ΔG = -27.2
1555 1576
ΔG = -24.6
1266 1294
ΔG = -18.0
376 392
ΔG = -16.4
947 975
ΔG = -13.4

Figure 19.

Sequencing strategy used for the EcoRI-6 fragment of _N. crassa_ mtDNA.

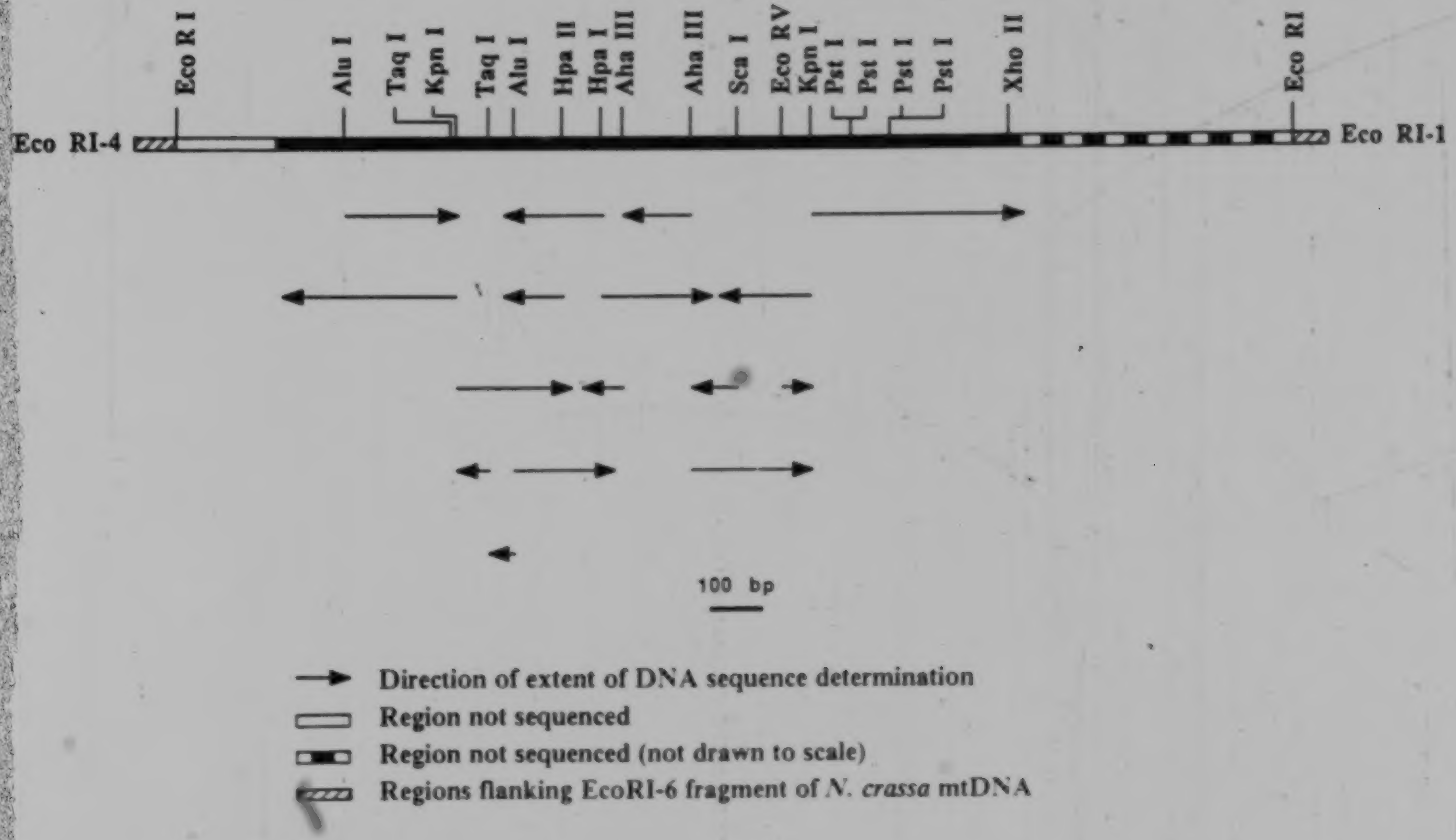

Eco R I
Alu I
Taq I
Kpn I
Taq I
Alu I
Hpa II
Hpa I
Aha III
Aha III
Sca I
Eco RV
Kpn I
Pst I
Pst I
Pst I
Pst I
Xho II
Eco RI
Eco RI-4
Eco RI-1
100 bp
Direction of extent of DNA sequence determination
Region not sequenced
Region not sequenced (not drawn to scale)
Regions flanking EcoRI-6 fragment of *N. crassa* mtDNA

Figure 20.

DNA sequence of a portion of the *N. crassa* EcoRI-6 mtDNA fragment.

Only the bottom strand was sequenced in the region from nucleotides 1 to 128. The sequence of the top strand was deduced from this information. Both strands were sequenced from nucleotide 129 to nucleotide 1070.
Only the top strand was sequenced from nucleotide 1071 to nucleotide 1477. The bottom strand sequence was deduced from this information. The first nucleotide of the sequence as shown was arbitrarily designated as position 1. The numbers on top of every tenth residue refer to nucleotide positions starting from position 1 as defined.
Nucleotides 1141 to 1154 inclusive form the first perfect PstI palindromic sequence and the nucleotides 1214 to 1227 inclusive form the second perfect PstI palindromic sequence.

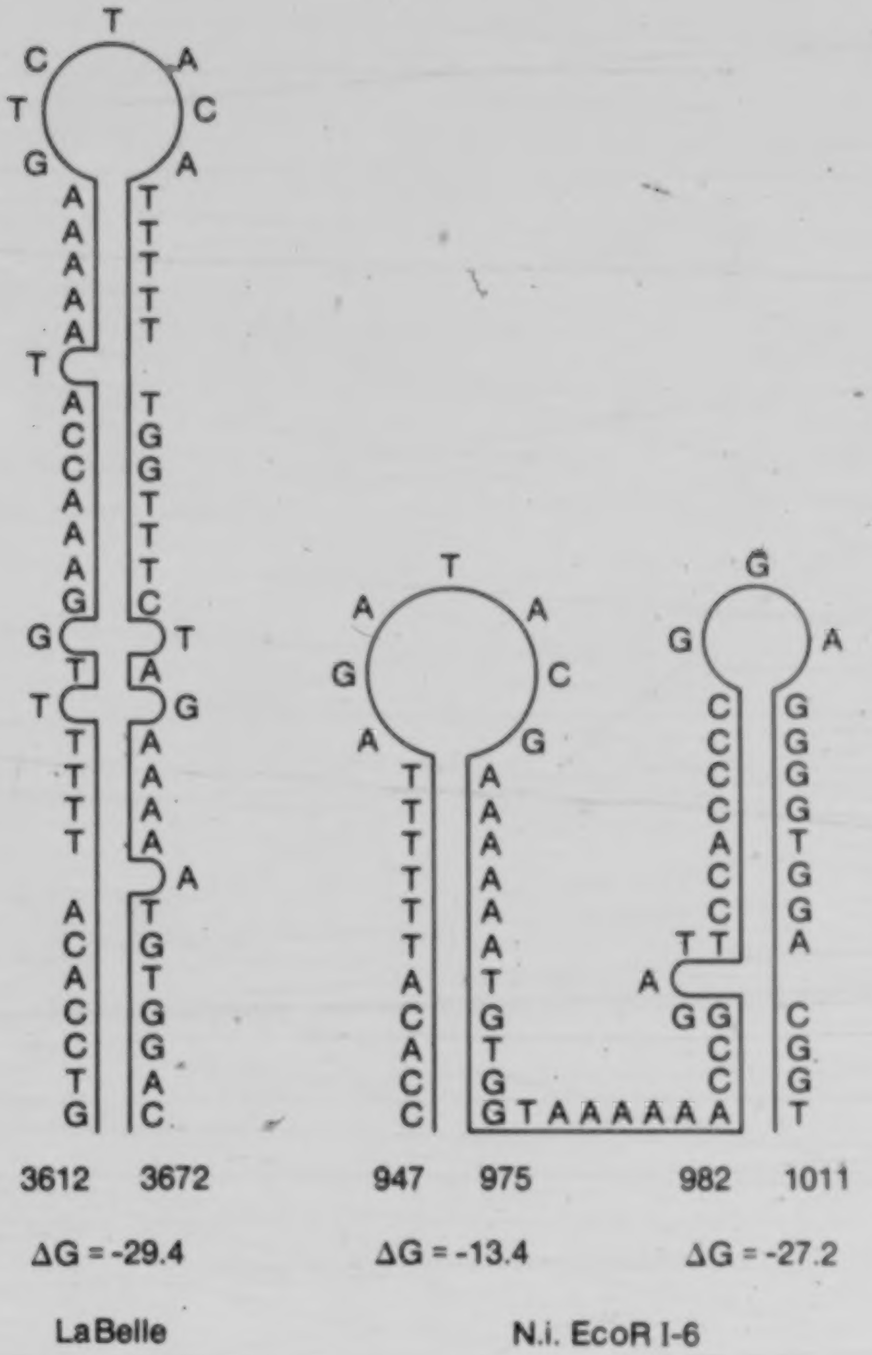
```
        10         20         30         40         50         60
GTCCCCCGAA AAATAAGGAT ACTAAAAGAA ACGGTAAATA CGCCTGTATT TGTAAAAGGA
CAGGGGGCTT TTTATTCCTA TGATTTTCTT TGCCATTTAT GCGGACATAA ACATTTTCCT

        70         80         90        100        110        120
CGTACGCTGT AGGGGGTGGA AAAGGTTATA AAAGTACTAT TGACCGGGGT GGATGCATTC
GCATGCGACA TCCCCCACCT TTTCCAATAT TTTCATGATA ACTGGCCCCA CCTACGTAAG

       130        140        150        160        170        180
ATTAAAGCAG CTTTCTCCGA TAGTGAACGT TTTGGTATAC CTGAAAATGA GTTTAAAACT
TAATTTCGTC GAAAGAGGCT ATCACTTGCA AAACCATATG GACTTTTACT CAAATTTTGA

       190        200        210        220        230        240
TATAATTCCA CTCCACTCCC CCCCCCCTTA AGGGGGGGGG AGTGGATTTT AGACGCCTTG
ATATTAAGGT GAGGTGAGGG GGGGGGGAAT TCCCCCCCCC TCACCTAAAA TCTGCGGAAC

       250        260        270        280        290        300
CGGCTCTGAC CAAGCAAAGA AAGTAAAACT TACAAAAACT AGTATTTCGG GGGACCCCGC
GCCGAGACTG GTTCGTTTCT TTCATTTTGA ATGTTTTTGA TCATAAAGCC CCCTGGGGCG

       310        320        330        340        350        360
TTCGGGGGCG GGGGCCCAAG TTAAAATGTC GTAAAACGAT ATCGAGAACG GTACCTCGGA
AAGCCCCCGC CCCCGGGTTC AATTTTACAG CATTTTGCTA TAGCTCTTGC CATGGAGCCT

       370        380        390        400        410        420
CAAAAGAAAC AGAAGGATTT ATAGAGTATG TAAAGAAACG AATAAGTACA TTCGATTAGA
GTTTTCTTTG TCTTCCTAAA TATCTCATAC ATTTCTTTGC TTATTCATGT AAGCTAATCT

       430        440        450        460        470        480
GCCGCAAGGC GTCTGTGAAT TAAATGTAAA TGTAAGCTAT TTTTTGATAA AATTGTTAGA
CGGCGTTCCG CAGACACTTA ATTTACATTT ACATTCGATA AAAAACTATT TTAACAATCT

       490        500        510        520        530        540
GGTGTATAAG CGCTCGTTTA GTTTGCTTTG CTTTGTGTGT AAAAAAGTCC ACATTTTTTA
CCACATATTC GCGAGCAAAT CAAACGAAAC GAAACACACA TTTTTTCAGG TGTAAAAAAT

       550        560        570        580        590        600
GATACGAAAA AATGTGGTAA AAAACCGGAT TCCACCCCGG AGGGGTGGAC GGTAACTGTG
CTATGCTTTT TTACACCATT TTTTGGCCTA AGGTGGGGCC TCCCCACCTG CCATTGACAC

       610        620        630        640        650        660
ACAATTATAG ACATGCGGGT GTGTGGGACC ATGACATTAC GTCAGGTTAA CGAAGTATCA
TGTTAATATC TGTACGCCCA CACACCCTGG TACTGTAATG CAGTCCAATT GCTTCATAGT

       670        680        690        700        710        720
GTAAAAATAA TAATAGTATG TTTAAATATT ATAATGTATG TGTGGTAGTA GTAATAATAA
CATTTTTATT ATTATCATAC AAATTTATAA TATTACATAC ACACCATCAT CATTATTATT

       730        740        750        760        770        780
TATTATTAAA TATTTAATTA TTTTTAACTT CGGGAATTTT TAATTATCAA ATAAGAATAT
ATAATAATTT ATAAATTAAT AAAAATTGAA GCCCTTAAAA ATTAATAGTT TATTCTTATA

       790        800        810        820        830        840
GAAGGATGCA ACGGTTTATT AGTGTAAGAA ATATTTTATT ATTTTAAAAT TAAAAAACAG
CTTCCTACGT TGCCAAATAA TCACATTCTT TATAAAATAA TAAAATTTTA ATTTTTTGTC

       850        860        870        880        890        900
TATGTTTACA ACCTGTGAGG TATCTCACAG GTGTGTTGTT GTTGTTGTTG TAGTTGTTGT
[illegible]
```

```
       910        920        930        940        950        960
TGCTTATAAG TACTAAGTGC TAACTGCTGC CTACTTAAAT AGTTTGGTAA AGTATTTATC
ACGAATATTC ATGATTCACG ATTGACGACG GATGAATTTA TCAAACCATT TCATAAATAG

       970        980        990       1000       1010       1020
CTCACATCGT AAGGTGTGGG TAGGATTAAA GTTTGAGTAT TGGGCGGATA TCCTCGCAAC
GAGTGTAGCA TTCCACACCC ATCCTAATTT CAAACTCATA ACCCGCCTAT AGGAGCGTTG

      1030       1040       1050       1060       1070       1080
GTTATTTATA ATCCATGTGA CGAATACCCA ACTCTCAAGG CATGGGTACC TGATAGAGTG
CAATAAATAT TAGGTACACT GCTTATGGGT TGAGAGTTCC GTACCCATGG ACTATCTCAC

      1090       1100       1110       1120       1130       1140
GGCCTATTTG TGCTTAGTAG GCATCGGCGT GAGTCTAAAT CGCAACCCCC TCCTTTCCCC
CCGGATAAAC ACGAATCATC CGTAGCCGCA CTCAGATTTA GCGTTGGGGG AGGAAAGGGG

      1150       1160       1170       1180       1190       1200
CTGCAGTACT GCAGGGGGGT GGGGAAGGGG AGGGGGAAGA GGAGGGGGGA TTTGCCTTTT
GACGTCATGA CGTCCCCCCA CCCCTTCCCC TCCCCCTTCT CCTCCCCCCT AAACGGAAAA

      1210       1220       1230       1240       1250       1260
TCGTTCCCCC GCCCTGCAGT ACTGCAGGGG GGGGTGTAAT AGAAATATTA TATAAGTGCT
AGCAAGGGGG CGGGACGTCA TGACGTCCCC CCCCACATTA TCTTTATAAT ATATTCACGA

      1270       1280       1290       1300       1310       1320
AAGAGGTAGG ACGGTTCGAA TCCGGAGGCT AAGAATATTT AAATTTAAAT CCCCCCCTCC
TTCTCCATCC TGCCAAGCTT AGGCCTCCGA TTCTTATAAA TTTAAATTTA GGGGGGGAGG

      1330       1340       1350       1360       1370       1380
TCTTACGTAT CCGAAGGGAG GCACGCCCCT CAAGGCCCGA GGGGTGGAGG AGGGGGTAGC
AGAATGCATA GGCTTCCCTC CGTGCGGGGA GTTCCGGGCT CCCCACCTCC TCCCCCATCG

      1390       1400       1410       1420       1430       1440
GAGGAGGGGT AAGGCTCCGC CTCAGCGAGG CTAAATTATT ATGATATTTG TTTACGTCAT
CTCCTCCCCA TTCCGAGGCG GAGTCGCTCC GATTTAATAA TACTATAAAC AAATGCAGTA

      1450       1460       1470
AATTGGATCT TAGTCATAAT TAAGTGCCGA GGTTGGG
TTAACCTAGA ATCAGTATTA ATTCACGGCT CCAACCC
```

Figure 21.

Potential hairpin loop structures in the region of homology between the LaBelle mitochondrial plasmid and the *N. intermedia* EcoRI-6 mtDNA fragment.

The numbering corresponds to Figs. 2 and 15 for LaBelle and *N. intermedia* EcoRI-6 respectively.
Unpaired regions are shown as loops or single stranded regions.
ΔG indicates the free energy of formation of these hairpin loop structures in Kcal/mol.


C T C T A C A
G A C A
A A A A A
T
A C C A A A G
G
T
T
T T T T
A C A C C T G
T T T T T
T G G T T T C
T
A
G
A A A A
A
T G T G G A C
3612
3672
ΔG = -29.4
LaBelle
A G A T A C G
T T T T T T A C A C C
A A A A A A T G T G G
T A A A A A A
G G A
C C C C A C C
T T
A
G G
G C C A
G G G G T G G A
C G G T
947
975
982
1011
ΔG = -13.4
ΔG = -27.2
N.i. EcoR I-6

PHOTOGRAPHIC PLATES

Plate I.

Northern blot showing hybridization of various probes to mtRNA isolated from the LaBelle plasmid-containing strain.

Lane 1 and 3 each contain 2.8 μg of mtRNA isolated from strain NCN47 (plasmid-containing strain).
Lane 2 and 4 each contain 3.9 μg of mtRNA isolated from strain NCN10 (non plasmid-containing strain).
Lanes 1 and 2 were probed with pBSA15 T3 riboprobe.
Lanes 3 and 4 were probed with pBSA15 T7 riboprobe.
The lower case letters a, b and c indicated on the side by longer arrows are the positions of rabbit reticulocyte and yeast rRNAs used as size markers.
"a" corresponds to 5.0 Kb rabbit reticulocyte 28S rRNA.
"b" corresponds to 3.75 Kb yeast 28S rRNA.
"c" corresponds to 2.0 Kb 18S rRNAs from both rabbit reticulocyte and yeast.
The numbers indicated on the side by shorter arrows are the estimated sizes of various bands in Kb.

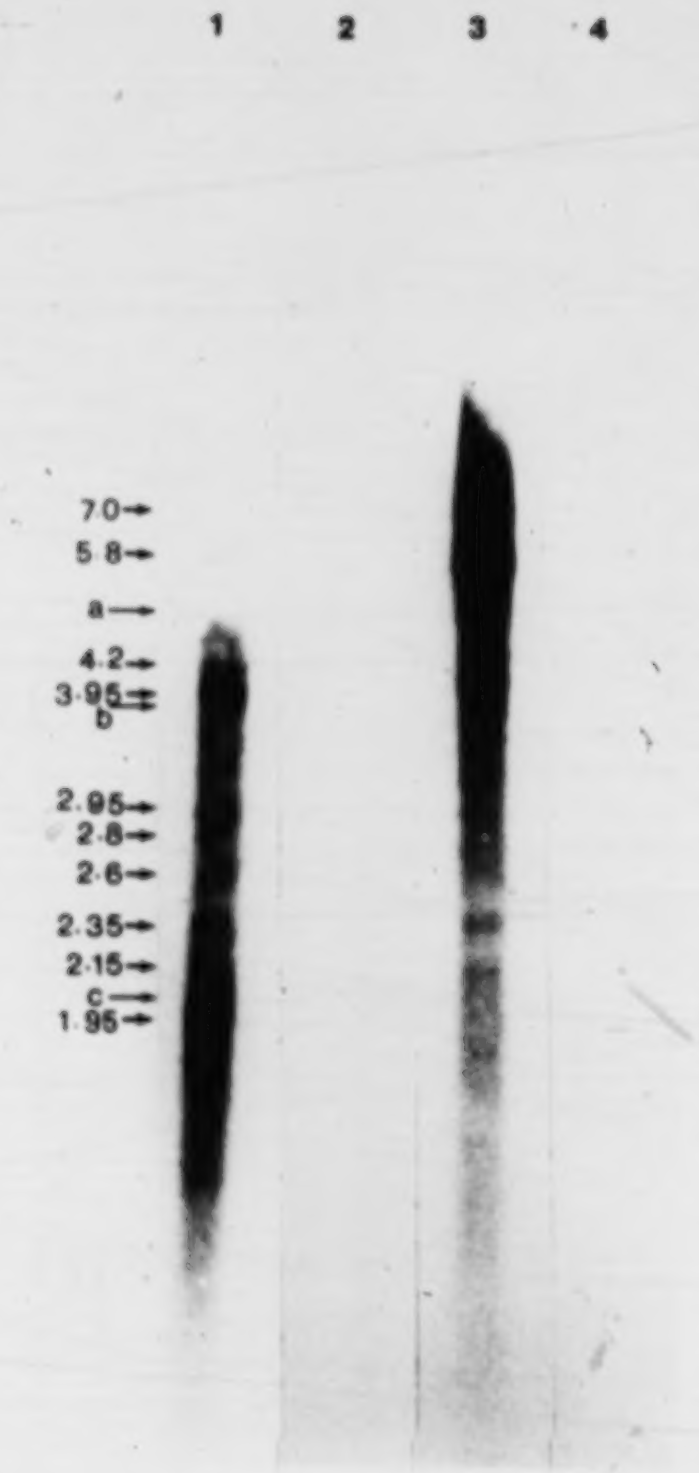

1
2
3
4
7.0
5.8
a
4.2
3.95
b
2.95
2.8
2.6
2.35
2.15
c
1.95

Plate II.

Ethidium bromide stained 1.5% agarose formaldehyde gel showing RNase A digestion is effective.

Lane 1 contains a sample (3.0 μg) of NCN47 mtRNA that was treated with DNase I (Pharmacia DNase I treated with Protease K in the presence of 10 mM $CaCl_2$ at 37°C for 1 hour) for 10 minutes at 37°C. The reaction was stopped by addition of DNase I stop buffer and the sample was phenol extracted followed by ethanol precipitation before loading on the gel.
Lane 2 contains a sample (3.0 μg) of NCN47 mtRNA that was treated with RNase A (Sigma) for 10 minutes at 37°C. The sample was phenol extracted and ethanol precipitated before loading on the gel.
Lane 3 contains 3.0 μg of NCN47 mtRNA.
Lane 4 contains 3.9 μg of NCN10 mtRNA.
Lane 5 contains 4 μg of yeast rRNA.
Lane 6 contains 4 μg of rabbit reticulocyte rRNA.
Arrows labelled 25S and 19S indicate the position of Neurospora mtrRNA bands.
The numbers on the side indicate sizes in Kb of size markers.

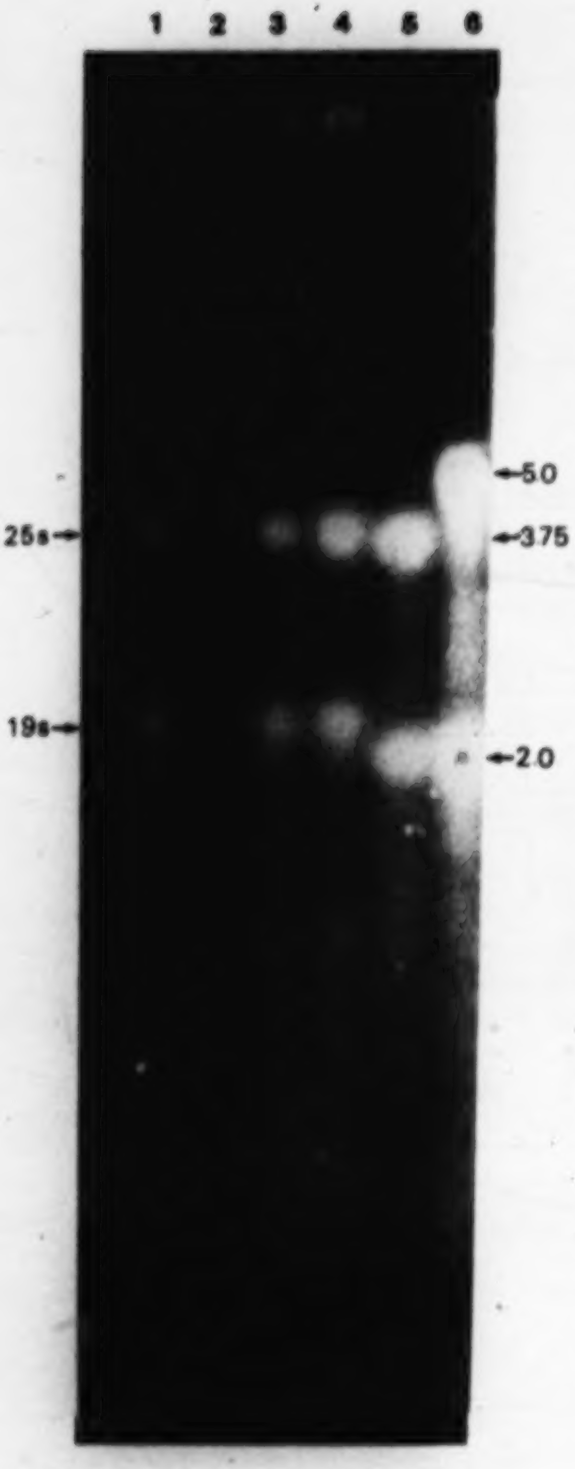

1
2
3
4
5
6
25s
19s
5.0
3.75
2.0

Plate III.

Northern blot performed to assess the quality of the mtRNA preparation.

Lane 1 contains 4.5 μg of NCN47 mtRNA.
Lane 2 contains 5.8 μg of NCN10 mtRNA.
Lane 3 contains a sample (5.8 μg) of NCN10 mtRNA that has been treated with DNase I for 10 minutes at 37°C. The sample was phenol extracted and ethanol precipitated before loading on the gel.
The blot was probed with radiolabelled pEL6 plasmid. pEL6 contains the HindIII-7c fragment of EcoRI-3 mtDNA, which contains the cytochrome c oxidase subunit I gene, cloned into the HindIII site of the pUC19 plasmid. 1 x 10^7 cpm were used for hybridization and the results shown were obtained after one hour exposure of the blot.
The estimated size of the COI transcript under the experimental conditions used was 6.5 Kb. Burger et al., (1982) had estimated its size to be 6.8 Kb using yeast rRNA as standard. Precise S1 mapping experiments put the size of the COI transcript at 5.0 Kb (Burger and Werner, 1985).

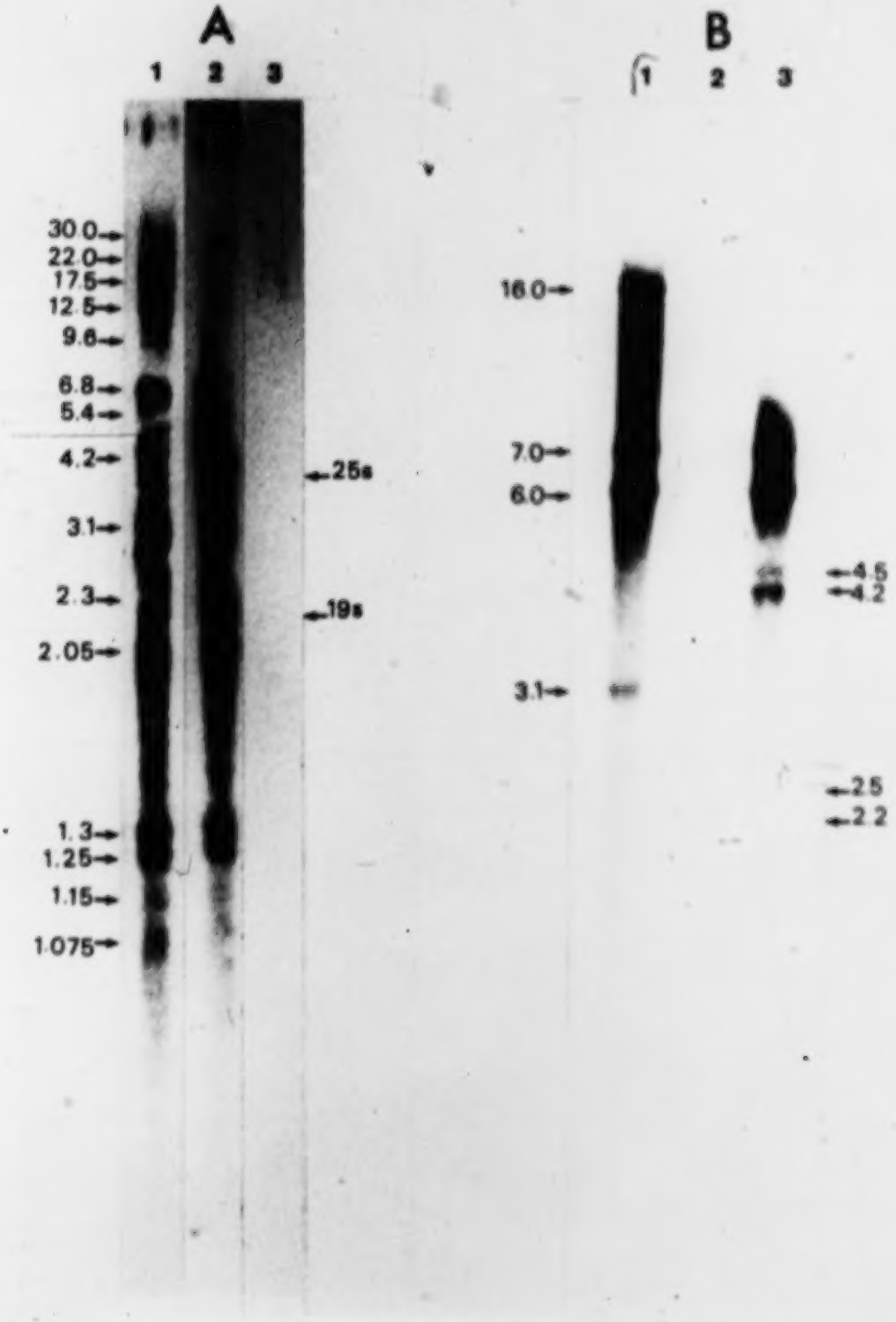
1
2
3
Col

Plate IV.

Northern blot showing LaBelle mitochondrial plasmid specific transcripts.

The blot in panel A was probed with 32.5 x 10^6 cpm pBSA15 T3 riboprobe and the results shown were obtained after 22 hours of exposure.
Lane 1 contains 3.0 μg of NCN47 mtRNA.
Lane 2 contains a sample (3.0 μg) of NCN47 mtRNA that was treated with DNase I (Pharmacia DNase I treated with Protease K in presence of 10 mM $CaCl_2$ at 37°C for 1 hour) for 10 minutes at 37°C. The reaction was stopped by addition of DNase I stop buffer and the sample was phenol extracted followed by ethanol precipitation before loading on the gel.
Lane 3 contains a sample (3.0 μg) of NCN47 mtRNA that was treated with RNase A (Sigma) for 10 minutes at 37°C. The sample was phenol extracted and ethanol precipitated before loading on the gel.

The blot in panel B was probed with 15.88 x 10^6 cpm pBSA15 T7 riboprobe and the results shown were obtained after 8 hours of exposure.
Lane 1 contains a sample (2.8 μg) of NCN47 mtRNA that was treated with RNase A+T1 (Sigma) for 10 minutes at 37°C. The sample was phenol extracted and ethanol precipitated before loading on the gel.
Lane 2 contains a sample (2.8 μg) of NCN47 mtRNA that was treated with DNase I (Promega biotech RQI) for 10 minutes at 37°C. The sample was phenol extracted and ethanol precipitated before loading on the gel.
Lane 3 contains 2.8 μg of NCN47 mtRNA.


A
1
2
3
30.0
22.0
17.5
12.5
9.6
6.8
5.4
4.2
3.1
2.3
2.05
1.3
1.25
1.15
1.075
25s
19s
B
1
2
3
16.0
7.0
6.0
3.1
4.5
4.2
2.5
2.2

BIBLIOGRAPHY

Adelman, J.P., Bond, C.T., Douglass, J., and Herbert, E. (1987). Two mammalian genes transcribed from opposite strands of the same DNA locus. Science *235* 1514-1517.

Akins, R.A., Grant, D.M., Stohl, L.L., Bottorff, D.A., Nargang, F.E., and Lambowitz, A.M. (Submitted). Mitochondrial plasmids of *Neurospora*. Nucleotide sequence of the Varkud plasmid and synthesis of a hybrid transcript with a 5' leader derived from mitochondrial RNA.

Akins, R.A., and Lambowitz, A.M. (1984). The [poky] mutant of *Neurospora* contains a 4 base-pair deletion at the 5' end of the mitochondrial small rRNA. Proc. Natl. Acad. Sci. USA *81*: 3791-3795.

Akins, R.A., and Lambowitz, A.M. (1985). General method for cloning *Neurospora crassa* nuclear genes by complementation of mutants. Mol. Cell. Biol. *5*: 2272-2278.

Akins, R.A., Kelley, R.L., and Lambowitz, A.M. (1986). Mitochondrial plasmids of Neurospora: Integration into mitochondrial DNA and evidence for reverse transcription in mitochondria. Cell *47*: 505-516.

Arends, H., and Sebald, W. (1984). Nucleotide sequence of the cloned mRNA and gene of the ADP/ATP carrier from *Neurospora crassa*. EMBO J. *3*: 377-382.

Baldacci, G., de Zamaroczy, M., and Bernardi, G. (1980). Excision sites in the GC clusters of the mitochondrial genome of yeast. FEBS Lett. *114*: 234-236.

Belasco, J.G., Beatty, J.T., Adams, C.W., von Gabain, A., and Cohen, S.N. (1985). Differential expression of photosynthesis genes in R. capsulata results from segmental differences in stability within the polycistronic *rxcA* transcript. Cell *40*: 171-181.

Bernard, U., Puhler, A., Mayer, F., and Kuntzel, H. (1975). Denaturation map of the circular mitochondrial genome of *Neurospora crassa*. Biochim. Biophys. Acta *402*: 270-278.

Bertrand, H. (1987). The *Kalilo* senescence factor of *Neurospora intermedia*: a mitochondrial IS-element derived from a nuclear plasmid. In Extrachromosomal elements in lower eukaryotes, R.B. Wickner, A. Hinnebusch, A.M. Lambowitz, I.C. Gunsalus, and A. Hollaender eds. (New York: Plenum Press). pp. 93-103.

Bertrand, H., and Pittenger, T.H. (1969). Cytoplasmic mutants selected from continuously growing cultures of *Neurospora crassa*. Genetics *61*: 643-659.

Bertrand, H., Chan, B.S.-S. and Griffiths, A.J.F. (1985). Insertion of a foreign nucleotide sequence into mitochondrial DNA causes senescence in *Neurospora intermedia*. Cell 41: 877-884.

Bertrand, H., Collins, R.A., Stohl, L.L., Goewert, R.R. and Lambowitz, A.M. (1980). Deletion mutants of *Neurospora crassa* mitochondrial DNA and their relationship to the "stop-start" growth phenotype. Proc. Natl. Acad. Sci. USA 77: 6032-6036.

Bertrand, H., Griffiths, A.J.F., Court, D.A. and Cheng, C.K. (1986). An extrachromosomal plasmid is the etiological precursor of KalDNA insertion sequences in the mitochondrial chromosome of senescent Neurospora. Cell 47: 829-837.

Birchmeier, C., Schumperli, D., Sconzo, G., and Birnstiel, M.L. (1984). 3' editing of mRNAs: Sequence requirements and involvement of a 60-nucleotide RNA in maturation of histone mRNA precursors. Proc. Natl. Acad. Sci. USA 81: 1057-1061.

Birnstiel, M.L., Busslinger, M., and Strub, K. (1985). Transcription termination and 3' processing: The end is in site! Cell 41: 349-359.

Breitenberger, C.A., Browning, K.S., Alzner-deWeerd, B., and RajBhandary, U.L. (1985). RNA processing in *Neurospora crassa* mitochondria: Use of transfer RNA sequences as signals. EMBO J. 4: 185-195.

Browning, K.S., and RajBhandary, U.L. (1982). Cytochrome oxidase subunit III gene in *Neurospora crassa* mitochondria. J. Biol. Chem. 257: 5253-5256.

Burger, G., Helmer-Citterich, M., Nelson, M.A., Werner, S., and Macino, G. (1985). RNA processing in *Neurospora crassa* mitochondria: Transfer RNAs punctuate a large precursor transcript. EMBO J. 4: 197-204.

Burger, G., Scriven, C., Machleidt, W., and Werner, S. (1982). Subunit 1 of cytochrome oxidase from *Neurospora crassa*: nucleotide sequence of the coding gene and partial amino acid sequence of the protein. EMBO J. 1: 1385-1391.

Burger, G., and Werner, S. (1985). The mitochondrial URF1 gene in *Neurospora crassa* has an intron that contains a novel type of URF. J. Mol. Biol. 186: 231-242.

Burger, G., and Werner, S. (1986). Mitochondrial gene URFN of *Neurospora crassa* codes for a long polypeptide with highly repetitive structure. J. Mol. Biol. 191: 589-599.

Burke, J.M., and RajBhandary, U.L. (1982). Intron within the large rRNA gene of N. crassa mitochondria: A Long Open Reading Frame and a consensus sequence possibly important in splicing. Cell 31: 509-520.

Burke, J.M., Belfort, M., Cech, T.R., Davies, R.W., Schweyen, R.J., Shub, D.A., Szostak, J.W., and Tabak, H.F. (1987). Structural conventions for group I introns. Nucl. Acids Res. 15: 7217-7221.

Burke, J.M., Breitenberger, C., Heckman, J.E., Dujon, B., and RajBhandary, U.L. (1984). Cytochrome b gene of Neurospora crassa mitchondria. J. Biol. Chem. 259: 504-511.

Campbell, A.M. (1969). Fertility factors. In Episomes : Modern Perspectives in Biology, H.O. Halvorson, H.L. Roman, and E. Bell eds. (New York: Harper & Row publishers) pp. 6-10.

Capaldi, R.A., and Vanderkooi, G. (1972). The low polarity of many membrane proteins. Proc. Natl. Acad. Sci. USA 69: 930-932.

Carignani, G., Groudinsky, O., Frezza, D., Schiavon, E., Bergantino, E. and Slonimski, P.P. (1983). An mRNA maturase is encoded by the first intron of the gene for the subunit I of the cytochrome oxidase in S. cerevisiae. Cell 35: 733-742.

Carpenter, C.D., Bruskin, A.M., Spain, L.M., Eldon, E.D., and Klein, W.H. (1982). The 3' untranslated regions of two related mRNAs contain an element highly repeated in the sea urchin genome. Nucl. Acids Res. 10: 7829-7842.

Chen, C.-n., Malone, T., Beckendorf, S.K., and Davis, R.L. (1987). At least two genes reside within a large intron of the dunce gene of Drosophila. Nature 329: 721-724.

Chung, S., Landfear, S.M., Blumberg, D.D., Cohen, N.S., and Lodish, H.F. (1981). Synthesis and stability of developmentally regulated dictyostelium mRNAs are affected by cell-cell contact and cAMP. Cell 24: 785-797.

Clark-Walker, G.D., McArthur, C.R., and Sriprakash, K.S. (1985). Location of transcriptional control signals and transfer RNA sequences in Torulopsis glabrata mitochondrial DNA. EMBO J. 4: 465-473.

Collins, R.A., and Lambowitz, A.M. (1981). Characterization of a variant Neurospora crassa mitochondrial DNA which contains tandem reiterations of a 1.9 Kb sequence. Curr. Genet. 4: 131-133.

Collins, R.A. and Lambowitz, A.M. (1983). Structural variations and optional introns in the mitochondrial DNAs of Neurospora strains isolated from nature. Plasmid 9: 53-70.

Collins, R.A., Stohl, L.L., Cole, M.D., and Lambowitz, A.M. (1981). Characterization of a novel plasmid DNA found in mitochondria of N. crassa. Cell *24*: 443-452.

Connaughton, J.F., Rairkar, A., Lockard, R.E., and Kumar, A. (1984). Primary structure of rabbit 18S ribosomal RNA determined by direct RNA sequence analysis. Nucl. Acids Res. *12*: 4731-4745.

Cosson, J., and Tzagoloff, A. (1979). Sequence homologies of (guanosine + cytidine) - rich regions of mitochondrial DNA of *Saccharomyces cerevisiae*. J. Biol. Chem. *254:* 42-43.

Cummings, D.J., Belcour, L. and Grandchamp, C. (1978). Etude au microscope electronique du DNA mitochondrial de *Podospora anserina* et presence d'une serie multimerique de molecules circulaires de DNA dans des cultures senescent. C.R. Acad. Sci. Paris *287*: 157-159.

Cummings, D.J., Turker, M.S., and Domenico, J.M. (1987). Mitochondrial excision-amplification plasmids in senescent and long-lived cultures of *Podospora anserina*. In Extrachromosomal elements in lower eukaryotes, R.B. Wickner, A. Hinnebusch, A.M. Lambowitz, I.C. Gunsalus, and A. Hollaender, eds. (New York: Plenum Press) pp. 129-146.

Davis, B.D. and Mingiola, E.S. (1950). Mutants of *Escherichia coli* requiring methionine or vitamin B12. J. Bact. *60*: 17-28.

Davis, R.H. and de Serres, F.J. (1970). Genetic and microbiological research techniques for *Neurospora crassa*. In Methods in Enzymology, H. Tabor, and C.W. Tabor, eds. (New York: Academic Press). *17A*: 79-143.

Dawson, A.J., Jones, V.P., and Leaver, C.J. (1984). The apocytochrome b gene in maize mitochondria does not contain introns and is preceded by a potential ribosome binding site. EMBO J. *3*: 2107-2113.

de Bruijn, M.H.L. (1983). *Drosophila melanogaster* mitochondrial DNA: A novel organization and genetic code. Nature *304*: 234-241.

de Jonge, J.C. and de Vries, H. (1983). The structure of the gene for subunit I of cytochrome c oxidase in *Neurospora crassa* mitochondria. Curr. Genet. *7*: 21-28.

de Vries, H., Alzner-DeWeerd, B., Breitenberger, C.A., Chang, D.D., de Jonge, J.C., and RajBhandary, U.L. (1986). The E35 stopper mutant of *Neurospora crassa*: Precise localization of deletion endpoints in mitochondrial DNA and evidence that the deleted DNA codes for a subunit of NADH dehydrogenase. EMBO J. *5*: 779-786.

de Vries, H., de Jonge, J. C., van't Sant, P., Agsteribbe, E., and Arnberg, A. (1981). A "Stopper" mutant of Neurospora crassa containing two populations of aberrant mitochondrial DNA. Curr. Genet. 3: 205-211.

Deno, H., Shinozaki, K., and Sugiura, M. (1984). Structure and transcription pattern of a tobacco chloroplast gene coding for subunit III of proton-translocating ATPase. Gene 32: 195-201.

Dujon, B., Colleaux, L., Jacquier, A., Michel, F., and Monteilhet, C. (1987). Mitochondrial introns as mobile genetic elements: The role of intron-encoded proteins. In Extrachromosomal elements in lower eukaryotes, R.B. Wickner, A. Hinnebusch, A.M. Lambowitz, I.C. Gunsalus, and A. Hollaender eds. (New York: Plenum Press). pp. 5-27.

Earl, P.L., Jones, E.V., and Moss, B. (1986). Homology between DNA polymerase of poxviruses, herpesviruses, and adenoviruses: Nucleotide sequence of the vaccinia virus DNA polymerase gene. Proc. Natl. Acad. Sci. USA 83: 3659-3663.

Esser, K., Kuck, U., Lang-Hinrichs, C., Lemke, P., Osiewacz, H.D., Stahl, U., and Tudzynski, P. (1986). Plasmids of Eukaryotes (New York: Springer-Verlag). pp. 7-55.

Esser, K., Kuck, U., Stahl, U., and Tudzynski, P. (1983). Cloning vectors of mitochondrial origin for eukaryotes: A new concept in genetic engineering. Curr. Genet. 7: 239-243.

Feinberg, A.P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132: 6-13.

Feinberg, A.P., and Vogelstein, B. (1984). Addendum of "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity." Anal. Biochem. 137: 266-267.

Fonty, G., Goursot, R., Wilkie, D., and Bernardi, G. (1978). The mitochondrial genome of wild-type yeast cells. VII. Recombination in crosses. J. Mol. Biol. 119: 213-235.

Foster, T.J. (1983). Plasmid-determined resistance to antimicrobial drugs and toxic metal ions in bacteria. Microbiol. Rev. 47: 361-409.

Francou, F. (1981). Isolation and characterization of a linear DNA molecule in the fungus Ascobolus immersus. Mol. Gen. Genet. 184: 440-444.

Gaillard, C., and Bernardi, G. (1979). The nucleotide sequence of the mitochondrial genome of a spontaneous "petite" mutant of yeast. Mol. Gen. Genet. 174: 335-337.

Gaillard, C., Strauss, F., and Bernardi, G. (1980). Excision sequences in the mitochondrial genome of yeast. Nature 283: 218-220.

Garber, R.C., Lin, J.J., and Yoder, O.C. (1987). Mitochondrial plasmids in *Cochliobolus heterostrophus*. In *Extrachromosomal elements in lower eukaryotes*. R.B. Wickner, A. Hinnebusch, A.M. Lambowitz, I.C. Gunsalus and A. Hollaender eds. (New York: Plenum Press). pp. 105-118.

Garber, R.C., Turgeon, B.G., and Yoder, O.C. (1984). A mitochondrial plasmid from the plant pathogenic fungus *Cochliobolus heterostrophus*. Mol. Gen. Genet. 196: 301-310.

Garfinkel, M.D., Pruitt, R.E., and Meyerowitz, E.M. (1983). DNA sequences, gene regulation and modular protein evolution in the *Drosophila* 68C glue gene cluster. J. Mol. Biol. 168: 765-789.

Garnier, J., Osguthorpe, D.J., and Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120: 97-120.

Griffiths, A.J.F., and Bertrand, H. (1984). Unstable cytoplasms in Hawaiian strains of *Neurospora intermedia*. Curr. Genet. 8: 387-398.

Griffiths, A.J.F., Kraus, S., and Bertrand, H. (1986). Expression of senescence in *Neurospora intermedia*. Can. J. Genet. Cytol. 28: 459-467.

Hager, K.M., Mandala, S.M., Davenport, J.W., Speicher, D.W., Benz, Jr. E.J., and Slayman, C.W. (1986). Amino acid sequence of the plasma membrane ATPase of *Neurospora crassa*: Deduction from genomic and cDNA sequence. *Proc. Natl. Acad. Sci*. USA 83: 7693-7697.

Hasson, J., Mougneau, E., Cuzin, F., and Yaniv, M. (1984). Simian virus 40 illegitimate recombination occurs near direct repeats. J. Mol. Biol. 177: 53-68.

Heckman, J.E., Sarnoff, J., Alzner-deWeerd, B., Yin, S., and RajBhandary, U.L. (1980). Novel features in the genetic code and codon reading patterns in *Neurospora crassa* mitochondria based on sequences of six mitochondrial tRNAs. *Proc*. Natl. Acad. Sci. USA 77: 3159-3163.

Henikoff, S., Keene, M.A., Fechtel, K., and Fristrom, J.W. (1986). Gene within a gene: Nested Drosophila genes encode unrelated proteins on opposite DNA strands. Cell 44: 33-42.

Hensgens, L.A.M., Bonen, L., de Haan, M., van der Horst, G., and Grivell, L.A. (1983). Two intron sequences in yeast mitochondrial COX1 gene: homology among URF-containing introns and strain-dependent variations in flanking exons. Cell 32: 379-389.

Hopp, T.P., and Woods, K.R. (1981). Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA 78: 3824-3828.

Hudspeth, M.E.S., Ainley, W.M., Shumard, D.S., Butow, R.A., and Grossman, L.I. (1982). Location and structure of the var1 gene on yeast mitochondrial DNA: nucleotide sequence of the 40.0 allele. Cell 30: 617-626.

Huiet, L., and Giles, N.H. (1986). The qa repressor gene of Neurospora crassa: Wild-type and mutant nucleotide sequences. Proc. Natl. Acad. Sci. USA 83: 3381-3385.

Jacquier, A., and Dujon, B. (1985). An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell 41: 383-394.

Jamet-Vierny, C., Begel, O., and Belcour, L. (1980). Senescence in Podospora anserina: Amplification of a mitochondrial DNA sequence. Cell 21: 189-194.

Kahn, M., Kolter, R., Thomas, C., Figurski, D., Meyer, R., Remaut, E., and Helinski, D.R. (1979). Plasmid cloning vehicles derived from plasmids Col E1, F, R6K, and RK2. In Methods in Enzymology, R. Wu, ed. (New York: Academic Press). 68: 268-280.

Kay, B.K., and Dawid, I.B. (1983). The 1723 element: A long , homogeneous, highly repeated DNA unit interspersed in the genome of Xenopus laevis. J. Mol. Biol. 170: 583-596.

King, P.V., and Blakesley, R.W. (1986). Optimizing DNA ligations for transformation. Focus 8: 1-3.

Kinnaird, J.H., and Fincham, J.R.S. (1983). The complete nucleotide sequence of the Neurospora crassa am (NADP - specific glutamate dehydrogenase) gene. Gene 26: 253-260.

Kinsey, J.A. (1985). Neurospora plasmids. In Gene Manipulations in Fungi, J.W. Benett, and L.L. Lasure eds. (New York: Academic Press). pp. 245-258.

Kistler, H.C., and Leong, S.A. (1986). Linear plasmid-like DNA in the plant pathogenic fungus Fusarium oxysporum f. sp. conglutinans. J. Bacteriol 167: 587-593.

Kotani, H., Ishizaki, Y., Hiraoka, H., and Obayashi, A. (1987). Nucleotide sequence and expression of the cloned gene of bacteriophage SP6 RNA polymerase Nucl. Acids Res. *15*: 2653-2664.

Kuiper, M.T.R., and de Vries, H. (1985). A recombinant plasmid carrying the mitochondrial plasmid sequence of *Neurospora intermedia* LaBelle yields new plasmid derivatives in *Neurospora crassa* transformants. Curr. Genet. *9*: 471-477.

Lambowitz , A.M. (1979). Preparation and analysis of mitochondrial ribosomes. In Methods in Enzymology, K. Moldave, and L. Grossman, eds.(New York: Academic Press). *59*: 421-433.

Lambowitz, A.M., Akins, R.A., Garriga, G., Henderson, M., Kubelik, A., and Maloney, K.A. (1985). Mitochondrial introns and mitochondrial plasmids of *Neurospora*. In Achievements and Perspectives in *Mitochondrial Research*, Volume II: Biogenesis, E. Quagliariello, E.C. Slater, F. Palmieri, C. Saccone, and A.M. Kroon eds. (Amsterdam: Elsevier Science Publishers). pp. 237-247.

Lambowitz, A.M., Akins, R.A., Kelley, R.L., Pande, S., and Nargang, F.E. (1987). Mitochondrial plasmids of *Neurospora* and other Filamentous Fungi. In Extrachromosomal *elements in* lower eukaryotes, R. Wickner, A. Hinnebusch, A. Lambowitz, I.C. Gunsalus, and A. Hollaender eds. (New York: Plenum Press). pp. 83-92.

Lang, B.F., Ahne, F., and Bonen, L. (1985). The mitochondrial genome of fission yeast *Schizosaccharomyces pombe*. The cytochrome b gene has an intron closely related to the first two introns in the *Saccharomyces cerevisiae cox 1* gene. J. Mol. Biol. *184*: 353-366.

Laughon, A., and Scott, M.P. (1984). Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. Nature *310*: 25-31.

Lazarus, C.M., Earl, A.J., Turner, G., and Kuntzel, H. (1980). Amplification of a mitochondrial DNA sequence in the cytoplasmically inherited "ragged" mutant of *Aspergillus amstelodami*. Eur. J. Biochem. *106*: 633-641.

Lazowska, J., Jacq, C., and Slonimski, P.P. (1980). Sequence of introns and flanking exons in wild type and box 3 mutants of cytochrome b reveals an interlaced splicing protein coded by an intron. Cell *22*: 333-348.

Legerton, T.L., and Yanofsky, C. (1985). Cloning and characterization of the multifunctional *his*-3 gene of *Neurospora crassa*. Gene *39*: 129-140.

Lennox, E.S. (1955). Transduction of linked genetic characters of the host by bacteriophage P1. Virology *1*: 190-206.

Locker, J., Lewin, A., and Rabinowitz, M. (1979). The structure and organization of mitochondrial DNA from petite yeast. Plasmid 2: 155-181.

Luiten, R.G.M., Putterman, D.G., Schoenmakers, J.G.G., Konings, R.N.H., and Day, L.A. (1985). Nucleotide sequence of the genome of PF3, an IncP-1 plasmid - specific filamentous bacteriophage of Pseudomonas aeruginosa J. Virol. 56: 268-276.

Macino, G., and Morelli, G. (1983). Cytochrome oxidase subunit 2 gene in Neurospora crassa mitochondria. J. Biol. Chem. 258: 13230-13235.

Macreadie, I.G., Scott, R.M., Zinn, A.R., and Butow, R.A. (1985). Transposition of an intron in yeast mitochondria requires a protein encoded by that intron. Cell 41: 395-402.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular cloning - A Laboratory Manual. (New York: Cold Spring Harbor Laboratory).

Masters, B.S., Stohl, L.L., and Clayton, D.A. (1987). Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. Cell 51: 89-99.

McGraw, N.J., Bailey, J.N., Cleaves, G.R., Dembinski, D.R., Gocke, C.R., Joliffe, L.K., MacWright, R.S., and McAllister, W.T. (1985). Sequence and analysis of the gene for bacteriophage T3 RNA polymerase. Nucl. Acids Res. 13: 6753-6766.

McKeown, M., and Firtel, R.A. (1981). Evidence for sub-families of Actin genes in Dictyostelium as determined by comparisons of 3' end sequences. J. Mol. Biol. 151: 593-606.

Messing, J. (1983) New M13 vectors for cloning. In Methods in Enzymology, R. Wu, L. Grossman, and K. Moldave eds. (New York: Academic Press) 101: 20-78.

Michel, F. and Dujon, B. (1983). Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast-, and nuclear-encoded members. EMBO J. 2: 33-38.

Michel, F., Jacquier, A., and Dujon, B. (1982). Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. Biochimie 64: 867-881.

Michel, F., and Lang, B.F. (1985). Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. Nature 316: 641-643.

Mills, D.R., and Kramer, F.R. (1979). Structure-independent nucleotide sequence analysis. Proc. Natl. Acad. Sci. USA 76: 2232-2235.

Moffatt, B.A., Dunn, J.J., and Studier, F.W., (1984). Nucleotide sequence of the gene for bacteriophage T7 RNA polymerase. J. Mol. Biol. 173: 265-269.

Mohan, M., Meyer, R.J., Anderson, J.B., and Horgen, P.A. (1984). Plasmid-like DNAs in the commercially important mushroom genus Agaricus. Curr. Genet. 8: 615-619.

Morelli, G., and Macino, G. (1984). Two intervening sequences in the ATPase subunit 6 gene of Neurospora crassa. J. Mol. Biol. 178: 491-507.

Mott, J.E., Galloway, J.L., and Platt, T. (1985). Maturation of Escherichia coli tryptophan operon mRNA: evidence for 3' exonucleolytic processing after rho-dependent termination. EMBO J. 4: 1887-1891.

Mylyk, O.M. (1975). Heterokaryon incompatibility genes in Neurospora crassa detected using duplication producing chromosome rearrangements. Genetics 80: 107-124

Nambu, J.R., Taussig, R., Mahon, A.C., and Scheller, R.H. (1983). Gene isolation with cDNA probes from identified Aplysia neurons: Neuropeptide modulators of cardiovascular physiology. Cell 35: 47-56.

Nargang, F.E. (1985). Fungal mitochondrial plasmids. Exper. Mycol. 9: 285-293.

Nargang, F.E. (1986). Conservation of a long open reading frame in two Neurospora mitochondrial plasmids. Mol. Biol. Evol. 3: 19-28.

Nargang, F.E., Bell, J.B., Stohl, L.L., and Lambowitz, A.M. (1983). A family of repetitive palindromic sequences found in Neurospora mitochondrial DNA is also found in a mitochondrial plasmid DNA. J. Biol. Chem. 258: 4257-4260.

Nargang, F.E., Bell, J.B., Stohl, L.L., and Lambowitz, A.M. (1984). The DNA sequence and genetic organization of a Neurospora mitochondrial plasmid suggest a relationship to introns and mobile elements. Cell 38: 441-453.

Nargang, F.E., and Bertrand, H. (1978). Nuclear mutants of Neurospora crassa temperature-sensitive for the synthesis of cytochrome aa_3. Mol. Gen. Genet. 166:15-23.

Natvig, D.O., May, G., and Taylor, J.W. (1984). Distribution and Evolutionary Significance of Mitochondrial plasmids in Neurospora spp. J. Bacteriol. 159: 288-293.

Nelson, M.A. and Macino, G. (1987). Structure and expression of the overlapping ND4L and ND5 genes of Neurospora crassa mitochondria. Mol. Gen. Genet. 206: 307-317

Netzker, R., Kochel, H.G., Basak, N., and Kuntzel, H. (1982). Nucleotide sequence of Aspergillus nidulans mitochondrial genes coding for ATPase subunit 6, cytochrome oxidase subunit 3, seven unidentified proteins, four tRNAs and L-rRNA. Nucl. Acids Res. 10: 4783-4794.

Newbury, S.F., Glazebrook. J.A., and Radford, A. (1986). Sequence analysis of the pyr-4 (orotidine 5'-P decarboxylase) gene of Neurospora crassa. Gene 43: 51-58.

Newbury, S.F., Smith, N.H., Robinson, E.C., Hiles, I.D., and Higgins, C.F. (1987). Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell 48: 297-310.

Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Kikyotani, S., Hirose, T., Asai, M., Takashima, H., Inayama, S., Miyata, T., and Numa, S. (1983). Primary structures of β - and δ - subunit precursors of Torpedo californica acetylcholine receptor deduced from cDNA sequences. Nature 301: 251-255.

Norgard, M.V., Keem, K., and Monahan, J.J. (1978). Factors affecting the transformation of Escherichia coli strain χ 1776 BY pBR322 Plasmid DNA. Gene 3: 279-292.

Norrañder, J., Kempe, T., and Messing, J. (1983). Construction of improved M13 vectors using oligodeoxy-nucleotide-directed mutagenesis. Gene 26: 101-106.

Ohtani, T., Uchimiya, H., Kato, A., Harada, H., Sugita, M., and Sugiura, M. (1984). Location and nucleotide sequence of a tobacco chloroplast DNA segment capable of replication in yeast. Mol. Gen. Genet. 195: 1-4.

Ojala, D., Montoya, J., and Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. Nature 290: 470-474.

Okita, T.W., Cheesbrough, V., and Reeves, C.D. (1985). Evolution and heterogeneity of the α-/β- type and γ- type gliadin DNA sequences. J. Biol. Chem. 260: 8203-8213.

Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G., and Steitz, T.A. (1985a). Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. Nature 313: 762-766.

Ollis, D.L., Kline, C., and Steitz, T.A. (1985b). Domain of E. coli DNA polymerase I showing sequence homology to T7 DNA polymerase. Nature 313: 818-819.

Osiewacz, H.D., and Esser, K. (1984). The mitochondrial plasmid of Podospora anserina : a mobile intron of a mitochondrial gene. Curr. Genet. 8: 299-305.

Ovchinnikov, Y.A. Monastyrskaya, G.S., Gubanov, V.V., Guryev, S.O., Salomatina, I.S., Shuvaeva, T.M., Lipkin, V.M., and Sverdlov, E.D. (1982). The primary structure of *E. coli* RNA polymerase. Nucleotide sequence of the rpoC gene and amino acid sequence of the β -subunit. Nucl. Acids Res. 10: 4035-4044.

Pabo, C.O., and Sauer, R.T. (1984). Protein DNA recognition. Ann. Rev. Biochem. 53: 293-321.

Posakony, J.W., Scheller, R.H., Anderson, D.M., Britten, R.J., and Davidson, E.H. (1981). Repetitive sequences of the sea urchin genome III. Nucleotide sequences of the cloned repeat elements. J. Mol. Biol. 149: 41-67.

Pring, D.R., Levings, C.S., III, Hu, W.W.L., and Timothy, D.H. (1977). Unique DNA associated with mitochondria in the "S"-type cytoplasm of male-sterile maize. Proc. Natl. Acad. Sci. USA 74: 2904-2908.

Pukui, M.K., and Elbert, S.H. (1957). Hawaiian-English Dictionary, University of Hawaii Press, Honolulu.

Queen, C., and Korn, L.J. (1984). A comprehensive sequence analysis program for the IBM Personal Computer. Nucl. Acids Res. 12: 581-599.

Rieck, A., Griffiths, A.J.F., and Bertrand, H. (1982). Mitochondrial variants of *Neurospora intermedia* from nature. Can. J. Genet. Cytol. 24: 741-759.

Roe, B.A., Ma, D.-P., Wilson, R.K., and Wong, J.F.-H. (1985). The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J. Biol. Chem. 260: 9759-9774.

Rosenberg, M., and Court, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. Ann. Rev. Genet. 13: 319-353.

Rubtsov, P.M., Musakhanov, M.M., Zakharyev, V.M., Krayev, A.S., Skryabin, K.G., and Bayev, A.A. (1980). The structure of the yeast ribosomal RNA genes. I. The complete nucleotide sequence of the 18S ribosomal RNA gene from Saccharomyces cerevisiae^{+}. Nucl. Acids Res. 8: 5779-5794.

Rutledge, B.J. (1984). Molecular characterization of the *qa*-4 gene of *Neurospora crassa*. Gene 32: 275-287.

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., and Petersen, G.B. (1982). Nucleotide sequence of bacteriophage λ DNA. J. Mol. Biol. 162: 729-773.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*: 5463-5467.

Schectman, M.G., and Yanofsky, C. (1983). Structure of the trifunctional *trp-1* gene from *Neurospora crassa* and its aberrant expression in *Escherichia coli*. J. Mol. Appl. Genet. *2*: 83-99.

Schinkel, A.H., Groot Koerkamp, M.J.A., Van der Horst, G.T.J., Touw, E.P.W., Osinga, K.A., Van der Bliek, A.M., Veeneman, G.H., Van Boom, J.H., and Tabak, H.F. (1986). Characterization of the promoter of the large ribosomal RNA gene in yeast mitochondria and separation of mitochondrial RNA polymerase into two different functional components. EMBO J. *5*: 1041-1047.

Schultz, G.E., Barry, C.D., Friedman, J., Chou, P.Y., Fasman, G.D., Finkelstein, A.V., Lim, V.I., Ptitsyn, O.B., Kabat, E.A., Wu, T.T., Levitt, M., Robson, B., Nagano, K. (1974). Comparison of predicted and experimentally determined secondary structure of adenyl kinase. Nature *250*: 140-142.

Schultz, G.E., and Schirmer, R.H. (1979). Principles of protein structure. (New York: Springer).

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., and Sugiura, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. *5*: 2043-2049.

Sinclair, J.H., Stevens, B.J., Sanghavi, P., and Rabinowitz, M. (1967). Mitochondrial-satellite and circular DNA filaments in yeast. Science *156*: 1234-1237.

Solymosy, F., Fedorcsak, I., Gulyas, A., Farkas, G.L., and Ehrenberg, L. (1968). A new method based on the use of diethyl pyrocarbonate as a nuclease inhibitor for the extraction of undegraded nucleic acid from plant tissues. European J. Biochem. *5*: 520-527.

Spradling, A., Hui, H., and Penman, S. (1975). Two very different components of messenger RNA in an insect cell line. Cell *4*: 131-137.

Staden, R. (1977). Sequence data handling by computer. Nucl. Acids Res. *4*: 4037-4051.

Staden, R. (1978). Further procedures for sequence analysis by computer. Nucl. Acids Res. *5*: 1013-1015.

Staden, R. (1979). A strategy of DNA sequencing employing computer programs. Nucl. Acids Res. *6*: 2601-2609.

Staden, R. (1980a). A computer program to search for tRNA genes. Nucl. Acids Res. *8*: 817-825.

Staden, R. (1980b). A new computer method for the storage and manipulation of DNA gel reading data. Nucl. Acids Res. *8*: 3673-3694.

Stahl, U., Lemke, P.A., Tudzynski, P., Kuck, U., and Esser, K. (1978). Evidence for plasmid like DNA in a filamentous fungus, the ascomycete *Podospora anserina*. Mol. Gen. Genet. *162*: 341-343.

Steinhilber, W., and Cummings, D.J. (1986). A DNA polymerase activity with characteristics of a reverse transcriptase in *Podospora anserina*. Curr. Genet. *10*: 389-392.

Stern, D.B., and Gruissem, W. (1987). Control of plastid gene expression: 3' inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell *51*: 1145-1157.

Stohl, L.L., Akins, R.A., and Lambowitz, A.M. (1984). Characterization of deletion derivatives of an autonomously replicating *Neurospora* plasmid. Nucl. Acids Res. *12*: 6169-6178.

Stohl, L.L., Collins, R.A., Cole, M.D., and Lambowitz, A.M. (1982). Characterization of two new plasmid DNAs found in mitochondria of wild type *Neurospora intermedia* strains. Nucl. Acids Res. *10*: 1439-1458.

Stohl, L.L., and Lambowitz, A.M. (1983). Construction of a shuttle vector for the filamentous fungus *Neurospora crassa*. Proc. Natl. Acad. Sci. USA *80*: 1058-1062.

Sugita, M., and Sugiura, M. (1984). Nucleotide sequence and transcription of the gene for the 32,000 dalton thylakoid membrane protein from *Nicotiana tabacum*. Mol. Gen. Genet. *195*: 308-313.

Sumner-Smith, M., Rafalski, J.A., Sugiyama, T., Stoll, M., and Soll, D. (1985). Conservation and variability of wheat α/β- gliadin genes. Nucl. Acids Res. *13*: 3905-3916.

Taylor, J.W., and May, G. (1987). *Neurospora* mitochondrial plasmids can be paternally inherited. Poster in Fungal Genetics Conference, Asilomar, April 1987.

Taylor, J.W., Smolich, B.D., and May, G. (1985). An evolutionary comparison of homologous mitochondrial plasmid DNAs from three *Neurospora* species. Mol. Gen. Genet. *201*: 161-167.

Taylor, J.W., Smolich, B.D., and May, G. (1986). Evolution and mitochondrial DNA in Neurospora crassa. Evolution 40: 716-739.

Terpstra, P., Holtrop, M., and Kroon, A. (1977). Heterogenous base distribution in mitochondrial DNA of Neurospora crassa. Nucl. Acids Res. 4: 129-139.

Thomas, P.S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA 77: 5201-5205.

Thomas, P.S. (1983). Hybridization of denatured RNA transferred or dotted to Nitrocellulose paper. In Methods in Enzymology, R. Wu, L. Grossman, and K. Moldave eds. (New York: Academic Press). 100: 255-266.

Tinoco, I. Jun., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M., and Gralla, J. (1973). Improved estimation of secondary structure in Ribonucleic Acids. Nature New Biol. 246: 40-41.

Tudzynski, P. (1982). DNA plasmids in eukaryotes with emphasis on mitochondria. Prog. Bot. 44: 297-307.

Tudzynski, P., Duvell, A., and Esser, K. (1983). Extrachromosomal genetics of Claviceps purpurea. I. Mitochondrial DNA and mitochondrial plasmids. Curr. Genet. 7: 145-150.

Tudzynski, P., and Esser, K. (1986). Extrachromosomal genetics of Calviceps purpurea. II. Plasmids in various wild strains and integrated plasmid sequences in mitochondrial genomic DNA. Curr. Genet. 10: 463-467.

Tullis, R.H., and Rubin, A. (1980). Calcium protects DNase I from proteinase K: A new method for the removal of contaminating RNase from DNase I. Anal. Biochem. 107: 260-264.

Viebrock, A., Perz, A. and Sebald, W. (1982). The imported proprotein proteolipid subunit of the mitochondrial ATP synthase from Neurospora crassa. Molecular cloning and sequencing of the mRNA. EMBO J. 1: 565-571.

Villa, V.D., and Storck, R. (1968). Nucleotide composition of nuclear and mitochondrial deoxyribonucleic acid of fungi. J. Bacteriol. 96: 184-190.

Vogel, H.J. (1964). Distribution of Lysine pathways among Fungi: Evolutionary implications. Am. Naturalist 98: 435-446.

Vollmer, S.J., and Yanofsky, C. (1986). Efficient cloning of genes of Neurospora crassa. Proc. Natl. Acad. Sci. USA 83: 4869-4873.

Waring, R.B., Davies, R.W., Scazzocchio, C., and Brown, T.A. (1982). Internal structure of a mitochondrial intron of *Aspergillus nidulans*. Proc. Natl. Acad. Sci. USA *79*: 6332-6336.

Wharton, K.A., Yedvobnick, B., Finnerty, V.G., and Artavanis-Tsakonas, S. (1985). *opa*: A novel family of transcribed repeats shared by the *notch locus* and other developmentally regulated loci in D. *melanogaster*. Cell *40*: 55-62.

Wildeman, A.G., Rasquinha, I., and Nazar, R.N. (1986). A "CAT" family of repetitive DNA sequences in *Saccharomyces cerevisiae*. J. Biol. Chem. *261*: 13401-13403.

Wong, H.C., and Chang, S. (1986). Identification of a positive retroregulator that stabilizes mRNAs in bacteria. Proc. Natl. Acad. Sci USA *83*: 3233-3237.

Wong, J.F.H., Ma, D.P., Wilson, R.K., and Roe, B.A. (1983). DNA sequence of the *Xenopus laevis* mitochondrial heavy and light strand replication origins and flanking tRNA genes. Nucl. Acids Res. *11*: 4977-4995.

Woudt, L.P., Pastink, A., Kempers-Veenstra, A.E., Jansen, A.E.M., Mager, W.H., and Planta, R.J. (1983). The genes coding for histone H3 and H4 in *Neurospora crassa* are unique and contain intervening sequences. Nucl. Acids Res. *11*: 5347-5360.

Wright, R.M., Horrum, M.A., and Cummings, D.J. (1982). Are mitochondrial structural genes selectively amplified during senescence in Podospora anserina? Cell *29*: 505-515.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene *33*: 103-119.

Yin, S., Burke, J., Chang, D.D., Browning, K.S., Heckman, J.E., Alzner-de Weerd, B., Potter, M.J., and RajBhandary, U.L. (1982). *Neurospora crassa* mitochondrial tRNAs and rRNAs: Structure, gene organization, and DNA sequences. In Mitochondrial Genes, P. Slonimski, P. Borst, and G. Attardi eds. (New York: Cold Spring Harbor Laboratory Press). pp. 361-373.

Yin, S., Heckman, J., and RajBhandary, U.L. (1981). Highly conserved GC-rich palindromic DNA sequences flank tRNA genes in Neurospora crassa mitochondria. Cell *26*: 325-332.

Zamaroczy, de. M., Faugeron-Fonty, G., Baldacci, G., Goursot, R., and Bernardi, G. (1984). The *ori* sequences of the mitochdonrial genome of a wild-type yeast strain: number, location, orientation and structure. Gene *32*: 439-457.